Absenteeism, 'Presenteeism', and Shirking

Sarah Brown and John G. Sessions[1]

ABSTRACT

We develop the Barmby et al. *(1993) model of absenteeism (which appeared in the* Scandanavian Journal of Economics*) to illustrate the potential for individuals to engage in 'presenteeism' viz. attending work even when they are entitled to paid absence.*

1. INTRODUCTION

ABSENCE REPRESENTS a major cost to society. The number of working days lost in the UK as a result of absence over the 1970s was at least as great as the number lost as a result of unemployment (Doherty, 1979).[2] In the year of the last miner's strike, 27 million working days were lost as a result of strike activity, a figure which pales by comparison with the 375 million working days lost on average as a result of absence over the 1980s (Brown and Sessions, 1996) The Confederation of British Industry has estimated that the total cost of absence to UK businesses in 2001 was over £12 billion, or one percent of GDP (CBI, 2002). For its part the Government has shown a keen awareness of the issue, its first-term position being that '... if it can achieve a 30 percent cut in absenteeism in public services by 2003, it would save taxpayers £1 billion a year.' (*The Economist*, 10 October 1998).

The costs of absence render it somewhat surprising that so many firms offer sick pay in excess of the statutory minimum. A survey of 1125 private and public sector employers found 803 offering company sick pay schemes as opposed to relying on the statutory sick pay scheme (see Brown, 1994). This apparent largesse is difficult to rationalize within the traditional model of absence in which workers endeavour to coincide actual and desired hours by 'rationally shirking' and providing less than required hours. Recently, however, a new perspective has emerged, highlighting the link between productivity and health, which recognizes that firms may find it preferable to discourage unhealthy/unproductive employees from working (see Chatterji and Tilley, 2002; Skåtun, 2003).[3]

In this paper we extend the Barmby *et al* (1993) — hereafter BST — model of absenteeism and illustrate the potential for individuals to engage in such 'presenteeism' i.e. to attend work even when they are entitled to paid absence. Such a situation may be problematic for firms, who may desire 'very sick' workers to remain at home to avoid the risk of costly epidemics and/or because their marginal product is too low, and who may therefore be acting optimally by offering 'excessive' sick-pay.

2. The model

We follow BST in modeling individuals as homogenous risk neutral utility maximisers endowed with a stock of time, T, which they allocate between work and leisure. Utility is an increasing function of income and leisure, with individuals attaching a weight to each depending upon some parameter, δ, representing their general level of health. δ is increasing in sickness and randomly distributed over the interval [0,1] with individuals valuing non-market (i.e. leisure) time more as $\delta \to 1$.[4]

$$U = (1-\delta)x + \delta l \tag{1}$$

where x denotes income and l denotes leisure. Prospective workers sign enforceable employment contracts that specify a wage, w, in return for a particular supply of effort. Considerations as to the intensity or quality of effort are ignored and for simplicity productivity is construed by mere attendance (i.e. h^c hours). There are no separation costs and firms are entitled to fire any worker not providing contracted labour hours. After the contract is signed, but before production commences, each worker realises his state of health and makes an ex post utility maximising decision as regards absence. There is no re-contracting and this decision depends on the outside options available.

There are two such options: first, all employees who absent themselves from work with 'acceptable' sickness are entitled to firm-financed sick pay, s. Sick pay is set exogenously and is payable to all employees for whom $\delta \geq \delta^z$, where δ^z is some exogenous minimum acceptable level of sickness. Second, all unemployed individuals are entitled to government financed unemployment insurance, b.

The assumption that sick pay exceeds unemployment insurance implies two critical levels of sickness relevant to individual decision making. First, an upper limit of sickness, δ^b, beyond which workers will never work and which is defined implicitly by $U^{na}(\delta^b) = U^u(\delta^b)$, where $U^{na}(U^u)$ denotes the expected utility of non-absence and $U^u(\delta^b)$ denotes expected utility from unemployment, such that:

$$\delta^b = \frac{w-b}{w-b+h^c} \tag{2}$$

And secondly, a level of sickness at which workers will be indifferent between non-absence and absence defined implicitly by $U^{na}\left(\delta^{s}\right)=U^{a}\left(\delta^{s}\right)$, where U^{a} denotes the expected utility of absence, such that:

$$\delta^{s}=\frac{w-s}{w-s+h^{c}} \tag{3}$$

BST assume that $\delta^{z}\in\left[\delta^{b},1\right]$.[5] Firms are required to pay s to all employees for whom $\delta\geq\delta^{z}$. Workers, however, prefer absence for all $\delta\geq\delta^{s}$ and given the asymmetry of information, have an incentive to overstate their true sickness for all $\delta\in\left[\delta^{s},\delta^{z}\right]$. Such 'shirking' is potentially costly to firms and may incite them to undertake monitoring. BST envisage a monitoring technology in which there is some probability, $\alpha < 1$, of each absentee's true state of health being revealed to the firm. Workers detected as shirking are fired immediately and are (only) able to claim unemployment insurance.

BST argue that the possibility of dismissal alters individual behavior as follows. In formulating their absence decision rule individuals will equate the expected utility of not absenting, $\mathbf{E}\left\{U^{na}\right\}=U^{na}$, with that of shirking, $\mathbf{E}\left\{U^{s}\right\}=\alpha U^{u}+\left(1-\alpha\right)U^{a}$, implying that individuals are indifferent between absence and non-absence at some 'reservation' sickness level, δ *, such that, $U^{na}\left(\delta^{*}\right)=\alpha U^{u}\left(\delta^{*}\right)+\left(1-\alpha\right)U^{a}$ where:

$$\delta^{*}=\frac{w-\beta}{w-\beta+h^{c}} \tag{4}$$

with $\beta=\alpha b+(1-\alpha)s$ and $\delta^{*}\in\left(\delta^{s},\delta^{b}\right)$. Note that the assumptions $s \leq b$ and ensures that $\delta^{z}\leq\delta^{*}$.

BST's decision rule is overly restrictive in assuming that individuals decide between shirking and non-shirking. Given that individuals form their reservation level of sickness before their health status is revealed to them, they will be unsure as to whether their actual level of sickness will exceed or fall short of the minimum general acceptable level of sickness. Thus it is arguably more appropriate, and certainly more general, to derive their reservation level of sickness through the equation of the expected utilities of absenting and not absenting:

$$\begin{aligned}&\mathrm{E}\left\{U^{na}\right\}=\mathrm{E}\left\{U^{a}\right\}\equiv\delta^{z}\mathrm{E}\left\{U^{s}\right\}+\left(1-\delta^{z}\right)\mathrm{E}\left\{U^{a}\right\}\\&\Rightarrow\\&U^{na}\left(\tilde{\delta}^{*}\right)=\delta^{z}\left[\alpha U^{u}\left(\tilde{\delta}^{*}\right)+\left(1-\alpha\right)U^{a}\left(\tilde{\delta}^{*}\right)\right]+\left(1-\delta^{z}\right)U^{a}\left(\tilde{\delta}^{*}\right)\end{aligned} \tag{5}$$

The right-hand-side of (5) denotes the expected utility to the worker of absent-

ing and is given by a weighted average of the expected utilities of valid and invalid (i.e. shirking) absence. Since sickness is uniformly distributed along the unit interval, there is a δ^z chance that the individual's actual level of sickness will be deemed invalid by the firm and therefore that any absence will be unauthorized. Conversely, there is a $(1\text{-}\delta^z)$ chance that the individual will be so sick that the firm would prefer the worker to absent and thus would authorize the absence. Solving (5) in terms of the worker's implied reservation level of sickness, $\tilde{\delta}^*$ implies:

$$\tilde{\delta}^* = \frac{w-\tilde{\beta}}{w-\tilde{\beta}+h^c} \tag{6}$$

where $\tilde{\beta} = \delta^z\beta + (1-\delta^z)s$. It is apparent from (5) and (6) that the reservation level of sickness is a function of wages, sick-pay, contractual hours, and the minimum acceptable level of sickness, δ^z, *viz* $\tilde{\delta}^* = \tilde{\delta}^*(\delta^z; w, s, h^c)$. The nature of the relationship between the reservation and authorized levels of sickness may be discerned from the derivatives and limits are set out in (7) and (8):

$$\frac{\partial\tilde{\delta}^*}{\partial\delta^z} = -\frac{\alpha(b-s)h^c}{\left(w-\tilde{\beta}+h^c\right)^2} > 0 \quad , \qquad \frac{\partial^2\tilde{\delta}^*}{\partial\left(\delta^z\right)^2} = -\frac{2\alpha^2(b-s)^2h^c}{\left(w-\tilde{\beta}+h^c\right)^3} < 0 \tag{7}$$

$$\lim_{\delta^z\to 1}\tilde{\delta}^* = \delta^* = \frac{w-\beta}{w-\beta+h^c} < 1 \quad , \qquad \lim_{\delta^z\to 0}\tilde{\delta}^* = \delta^s = \frac{w-s}{w-s+h} > 0 \tag{8}$$

It is apparent from (7) and (8) that the reservation level of sickness is concavely increasing in the minimum acceptable level of sickness. As illustrated in figure 1, however, the former's support lies within the unit interval implying that there are ranges of sickness over which the worker will always absent or attend irrespective of the acceptable level of sickness.

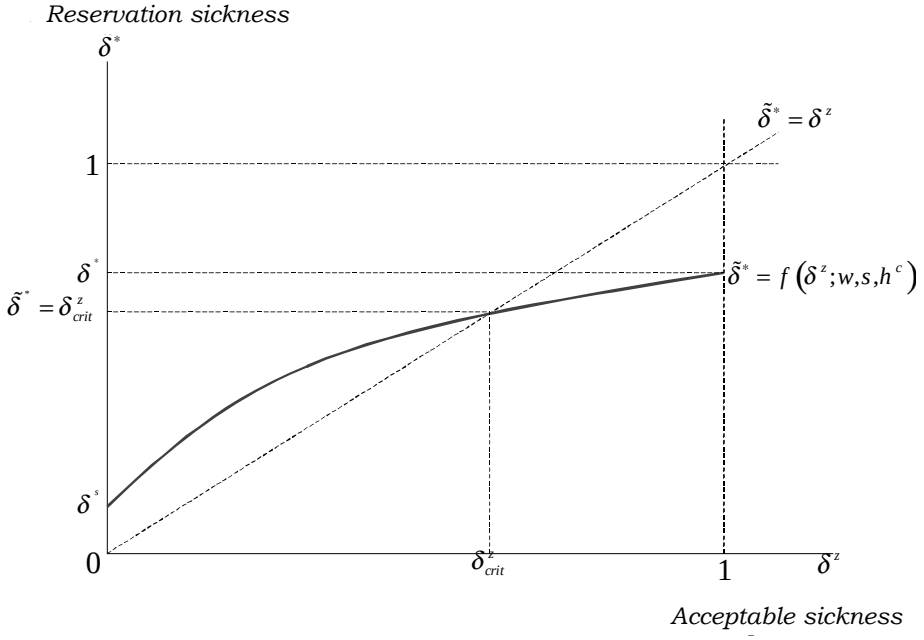
Indeed, we can define a critical level of acceptable sickness, δ^z_{crit}, at which $\tilde{\delta}^*(\delta^z_{crit}; w, s, h^c) = \delta^z_{crit}$ such that $\delta^z \le \delta^z_{crit} \Rightarrow \tilde{\delta}^* \ge \delta^z$ and $\delta^z > \delta^z_{crit} \Rightarrow \tilde{\delta}^* < \delta^z$ — setting the acceptable level of sickness below the critical level induces a reservation level of sickness above the acceptable level of sickness, and vice versa. This critical level is equal to:

$$\delta^z_{crit} = \frac{\left(w-s+h^c\right)-\alpha(s-b)\pm\sqrt{\left[\left(w-s+h^c\right)-\alpha(s-b)\right]^2+4\alpha(w-s)(s-b)}}{-2\alpha(s-b)} \tag{9}$$

Setting the acceptable level of absence below this critical level implies a reservation level of sickness above the acceptable level such that individuals may engage in presenteeism, attending work when the firm would prefer them

to absent. Conversely, setting the acceptable level of absence in excess of the critical rate implies a reservation level of sickness above the acceptable level such that individuals may be tempted to take illegitimate (i.e. shirking) absence.

Figure 1: Acceptable (δ^z) and Reservation ($\tilde{\delta}^*$) Levels of Sickness



To be sure:

$$\delta^z \le \delta^z_{crit} \Rightarrow \tilde{\delta}^* \ge \delta^z \Rightarrow \begin{cases} \textit{presenteeism over range} \quad (\delta^z, \tilde{\delta}^*) \\ \textit{valid absence over range} \quad (\tilde{\delta}^*, 1) \end{cases} \tag{10}$$

$$\delta^z > \delta^z_{crit} \Rightarrow \tilde{\delta}^* < \delta^z \Rightarrow \begin{cases} \textit{shirking absence over range} \quad (\tilde{\delta}^*, \delta^z) \\ \textit{valid absence over range} \quad (\delta^z, 1) \end{cases} \tag{11}$$

4. Concluding comments

Without specifying further the nature of the firm's production and/or monitoring technologies, it is difficult to comment precisely on the welfare implications. If shirking absence were unequivocally costly to the firm, in terms of lost output and prohibitive monitoring costs, then the firm would need to ensure that $\delta^z \le \delta^z_{crit}$. Conversely, if presenteeism so impinges upon output, either

through lost individual output or through the potential spread of costly epidemics, then the firm would need to ensure that $\delta^z > \delta^z_{crit}$. The key point from the present analysis is that the setting of δ^z is critical — too high and individuals will be tempted to shirk; too low and they may choose to attend work when their health suggests they should absent themselves.

The firm therefore needs to take care when setting the authorized absence level, both because it impinges upon actual absence, and because low absence is not always preferable. As *Labour Research* (1998) concludes, '... the lowest possible absence rates are not necessarily the best outcome for a firm or organization, as this may result in ... people coming to work when they should be at home, working below par, putting themselves at increased risk, passing on illnesses to their workmates, and undermining morale at work' (p.22). The cost of such behaviour to society has yet to be determined.

Accepted for publication: 6th October 2003

ENDNOTES

1. Brown: Department of Economics, University of Leicester, University Road, Leicester, Leicestershire LE1 7RH, England, email: sb118@leicester.ac.uk. Sessions: Department of Economics and International Development, University of Bath, Claverton Down, Bath, BA2 7AY, England, email. j.g.sessions@bath.ac.uk.

2. Similar evidence for Canada is presented by Akyeampong (1988). Evidence for the US indicates that in the late 1970's approximately 5 million working days per month were being lost as a result of worker absenteeism, equivalent to a loss of more than $2 billion in wages and salaries per month (Dunn and Youngblood, 1986).

3. This approach is not entirely without antecedence. In one of the pioneering studies of absence it was noted that '...some absenteeism may in fact be healthy for organizations in that such behavior can allow for temporary escape from stressful situations ... rigid attempts to ensure perfect attendance may lead to unintended and detrimental consequences on the job.' (Steers and Rhodes, 1978, p. 34).

4. This could be for the purpose of recuperation or because it becomes increasingly onerous to supply effort at higher levels of sickness.

5. Although this assumption is made primarily to simplify the exposition of the model some rationale can be given by supposing that firms prefer 'very' sick workers not to work in order to minimise the danger of costly epidemics or accidents. Moreover, such a assumption is reasonable given that the government is setting both parameters.

REFERENCES

Akyeampong E B (1988) 'Time Lost for Work for Personal Reasons', *The Labour Force*, Labour and Household Surveys Analysis Division, (Statistics Canada), May, 87-121.

Barmby T A, Sessions J G and Treble J G (1993) 'Absenteeism, Efficiency Wages and Shirking', *Scandinavian Journal of Economics*, 96 (4), 561-66.

Brown S (1994) 'Dynamic Implications of Absence Behavior', *Applied Economics,* 26, 1163-1175.

Brown S and Sessions J G (1996) 'The Economics of Absence: Theory and Evidence', *Journal of Economic Surveys,* 10, 23-53.

Chatterji M and Tilley C (2002) 'Sickness, Absenteeism, Presenteeism, and Sick Pay', *Oxford Economic Papers*, 54 (4), 669-87.

Confederation of British Industry (2002) 'Counting the Costs: 2002 CBI Absence and Labour Turnover Survey', London: CBI Publications Unit.

Doherty N A (1979) `National Insurance and Absence from Work.' *Economic Journal*, 89, 50-65.

Dunn L F and Youngblood S A (1986) 'Absenteeism as a Mechanism for Approaching an Optimal Labour Market Equilibrium: An Empirical Study', *Review of Economics and Statistics*, 68, 668-74.

Labour Research (1998) 'Noticeable by their Absence', August, 21-22.

Skåtun J D (2003) 'Take Some Days off Why Don't You? Endogenous Sick Leave and Pay', *Journal of Health Economics*, 22, 379-402.

Steers R M and Rhodes S R (1978) 'Major Influences on Employee Attendance: A Process Model', *Journal of Applied Psychology,* 63, 391-407.