University of Dundee

Sickness, absenteeism, "presenteeism" and sick pay

Chatterji, Monojit; Tilley, Colin J.

Publication date:
2000

Link to publication in Discovery Research Portal

Dundee Discussion Papers in Economics

Sickness, Absenteeism, "Presenteeism" and Sick Pay

Monojit Chatterji

and

Colin J. Tilley

Sickness, Absenteeism, 'Presenteeism' and Sick Pay

by

Monojit Chatterji

and

Colin J. Tilley*

University of Dundee
Dundee DD1 4HN

December 2000

Abstract: The annual cost of absenteeism from the workplace in the UK has been estimated to be over 1% of GDP. The traditional approach to a discussion of absence has been for the firm to passively accept both wages and sick pay and allow workers to choose their absence behaviour. Most empirical research has been based on this approach. However, if absence is costly why should firms pay extra-statutory sick pay? One reason may be the phenomena of presenteeism (ill workers attending work). This may adversely affect productivity. This paper shows that allowing for presenteeism has important implications for both the design of optimal wage-sick pay contracts and for the interpretation of empirical studies. Specifically, we show that firms will offer a level of sick pay greater than the statutory minimum.

<center>**Sickness, Absenteeism, "Presenteeism" and Sick Pay**</center>

Section 1: Introduction

Labour productivity is one of the major determinants of national prosperity in the long run. It is a commonplace observation that one of the determinants of the productivity of the labour force taken as a whole is the average intensity of its utilisation. Unemployment, strikes and absence from work can be thought of as three independent measures of utilisation of the work force. A rise in any of the three would appear to lower the productivity and output of the labour force taken as a whole.

There has been considerable research both theoretical and empirical on the factors determining unemployment and strikes. Policy analysis designed to reduce unemployment and strike activity has also been much discussed. By contrast, the analysis of the causes and incidence of absence has received less attention, until fairly recently. This is somewhat surprising since the incidence of absence as an empirical phenomena is at least as important as unemployment and strikes. Thus Brown (1994, p. 1163) reports that, 'the number of working days lost in the UK as a result of absence over the 1970s was at least as large as the number lost as a result of unemployment. In the year of the last miners' strike 27 million 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.'

Recently there has been an upsurge of public interest in the incidence of absence and its potentially damaging consequences on output. The Economist (10th October 1998) reports that, "..the government reckons that if it can achieve a 30% cut in absenteeism in public services by 2003, it would save taxpayers £1 billion a year." Similarly, the

<center>1</center>

CBI estimates that the total cost of absence to UK businesses in 1996 could have been as large as £12 billion.[1]

The traditional approach to explaining absence exploits the conventional labour-leisure choice model. Desired hours of work are determined from the workers indifference curves and the budget constraint s/he faces. However, official hours of work are set by employers or by convention. The difference between official hours required and hours of work desired by workers at the going wage provides the motivation for absence[2]. Thus absence is seen as mainly an attempt by workers to bring actual hours worked more into line with desired hours by effectively working less than the official hours required by the job. This "rational shirking" approach to absence has been exploited both theoretically and empirically by Allen (1981), Dunn and Youngblood (1986), Barmby and Treble (1991). More recently this conventional approach has been augmented by Brown (1994) (using a two period analysis), and Kahana and Weiss (1992) who develop a game theoretic model to compare the degree of absence in labour managed and profit maximising firms. All these studies have the common feature that the firm behaves passively whilst workers respond to a given wage-sick pay structure using their informational advantage. Barmby et al. (1994) make a departure from this tradition by assuming the firm sets the wage (in an efficiency wage manner) to try and control absence. Brown and Sessions (1996) provide a comprehensive review of this literature.

The explanation of absence is further complicated by the existence of sick pay. Many firms offer workers sick pay in excess of the statutory minimum. For example, Brown (1994) and Barmby et al. (1991) employ data from a single firm that pays sick pay at

three different levels. All are above the statutory minimum. Moreover, Brown (1994, p. 1168) reports that, '..[in] a survey of 1125 private and public sector employers 803 offered company sick pay schemes as opposed to relying on the statutory sick pay scheme.'.

The traditional model cannot account for this - and does not seek to do so. Indeed there is a fundamental tension between the basic tenet of the traditional approach (workers seeking to absent themselves not because they are sick but because they are being made to work longer hours than they want) and the observation that firms offer extra statutory sick pay. By offering sick pay, firms are in effect increasing the workers' incentives to absent themselves even more, thus potentially driving a greater wedge between actual hours worked and official hours set by the firm. What motivates firms to act in such an apparently irrational manner?

In a recent study Coles and Treble (1993) explicitly include sick pay as one of the control variables of the firm. Their approach still relies on a form of "rational shirking" in a principal-agent setting in which firms offer workers a contract that specifies wages and sick pay. Sickness is a random phenomenon and all sick workers are assumed to absent themselves and receive sick pay. Healthy workers have a choice as to whether or not to absent themselves. The contract obviously must satisfy the usual incentive compatibility conditions. The twist in this approach is that whilst the random event (sickness) cannot be observed, the relevant action (absence or presence) can be. Hence sickness pay - a form of contingent payment - is perfectly feasible.

Though this new approach has widened the analysis considerably, it still sees, implicitly at least, absence as undesirable from the firm's point of view. By assumption, workers are either healthy or sick and only healthy workers face any

choice problem at all. In designing the contract the firm is seeking to reduce absence of healthy workers as much as possible. However Coles and Treble (op. cit., 1993) do not investigate "presenteeism" - the flip side of absenteeism. Just as "healthy" workers absenting themselves is damaging, so might the situation in which "unhealthy" workers turn up for work if their productivity is adversely affected by their health state. This possibility and the potential impact it may have on the design of the optimal wage-sickness pay contract is one the main insights of this paper. Another advantage of our approach is that it suggests a modified interpretation of many of the empirical studies that posit a negative relationship between wages and absence (Dunn and Youngblood, 1986; Barmby and Treble, 1991; Drago and Wooden, 1992). Specifically, our analysis suggests that these studies may not be particularly robust.

The rest of the paper is organised as follows. In section 2 we briefly outline the conventional labour-leisure model of absence, the 'efficiency wage' model and the contract model of absence. Section 3 specifies our approach, highlighting the possibility of "presenteeism". In Section 4 we discuss the empirical implications of our results. Section 5 concludes the paper.

Section 2: Absence, Wages and Sickness Pay

2(i) The Supply Side Approach

The following section provides an explanation of absence within a standard labour-leisure choice model. The temptation to be absent is determined by the difference between the marginal rate of substitution between consumption and leisure and the real wage rate. This arises because official hours of work are assumed to be pre-determined.

If the individual is free to select his utility maximising choice of leisure hours[3] then an interior solution implies equality between the marginal rate of substitution and the real wage[4] and the temptation to absent would not arise. However, it is likely that the firm is not indifferent to when and for how long the individual attends[5]. This rationing of workers' labour supply choices is precisely what gives workers an incentive to move towards their optimal choice of hours by absenting themselves when 'involuntarily overemployed'[6]. Figure 1 illustrates a worker's choice diagrammatically.

Figure 1: The optimal choice of hours[7].

If the worker's choice of labour supply is unconstrained the optimal choice of hours worked is $\overline{h}$. However if hours are constrained at h^c where $MRS^c > w/p$ (like point A in figure 1) the agent has an incentive to adjust towards the optimal choice of leisure. The decision to take more leisure then implies some amount of absence[8,9]. The extent to which an individual can move towards his desired hours will, naturally, depend upon how strictly an employer enforces the terms of the contract and the nature of the budget set[10].

The introduction of a payment made by the firm when the individual is absent, which we term sick pay[11], at some rate s ($<w$), serves simply to alter the budget constraint the individual faces. Hence an individual who is absent is paid for fully contracted hours, h^c, but at rate s. The new budget constraint is represented by AB' in figure 2 below and has slope $(w - s)$.
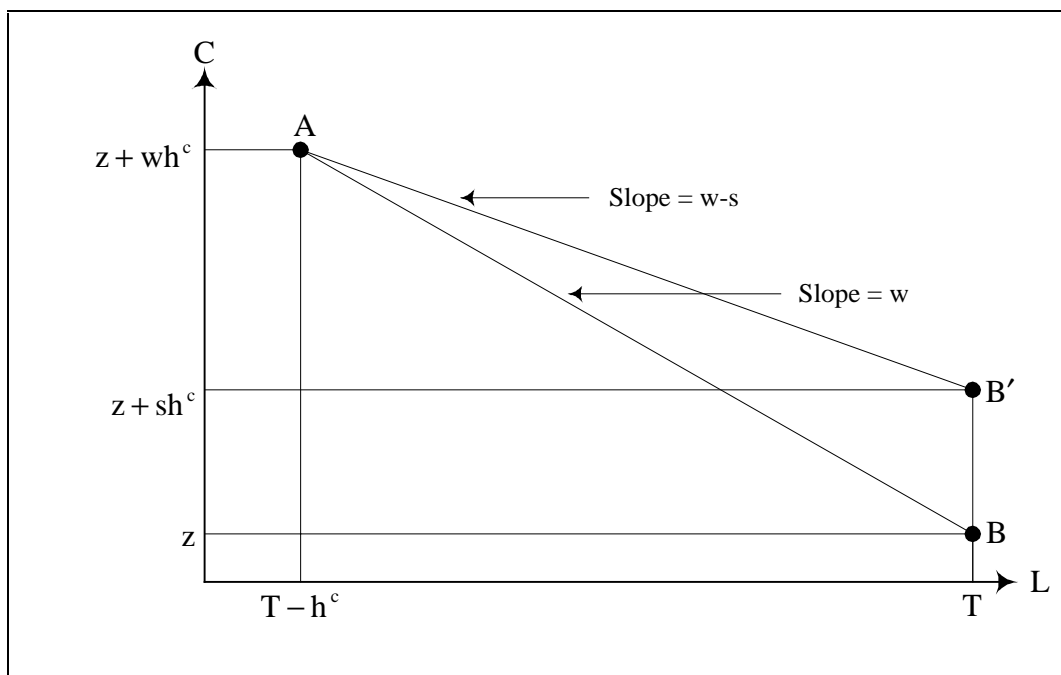

Figure 2: The introduction of sick pay.

As the rate of sick pay, s, rises point B increases (to B') and the slope of the budget constraint falls. Intuitively, as s increases the substitution effect predicts less labour

supply (the effective price of leisure has fallen) and the income effect reinforces the substitution effect if leisure is a normal good. Clearly, as the level of sick pay tends to a replacement ratio of one (AB' becomes horizontal), rational individuals will absent themselves completely.

Given the introduction of sick pay into the analysis, it seems natural to also introduce some index of health[12]. Denote this health index by θ where lower values of θ are associated with healthier states. It is assumed[13] that utility and the marginal utility of consumption are decreasing in ill health, i.e. $U_\theta < 0$ and $U_{C\theta} < 0$ [14]. As θ rises, the agent places more value on leisure relative to consumption and his indifference curves become steeper. This rise in the marginal rate of substitution increases the likelihood of absence[15]. The most general form of the labour-leisure choice model includes wages, sick pay and health state. The empirical implications of this model of absence can thus be stated as follows

$$\text{Absence} = f\left(\overset{-}{w} - s, \overset{+}{z}, \overset{+}{h^c}, \overset{+}{\theta} \right)[16].$$

While the 'supply side' approach has obvious advantages in its explanation of absence, it also has many drawbacks. Barmby et al. (1991) note that empirical studies of absence may not be robust given this approach's failure to consider the demand side. Two other drawbacks of the approach are of concern in this paper. Firstly, firms are assumed to have no discretion in the choice of the (w, s) contract and take these values as parametrically given. Secondly, the difficulty of monitoring absent workers gives rise to the possibility of strategic interaction between the worker and the firm: workers may claim sick pay when they are in good health. Presumably when firms are attempting to reduce absenteeism, this is precisely the sort of absence they wish to target. Absence arises essentially because of an unexplained restriction on the hours of

workers. In general terms, the usefulness of the 'supply side' approach is that it treats the firm as passive. This fundamental drawback of treating the firm as passive is addressed in different ways by the efficiency wage and contract models.

2(ii) The Efficiency Wage Approach

Barmby et al. (1994) recognise that individuals may absent without good cause[17]. If there is a distribution of health states then some discrepancy may arise between the health state at which the firm will pay (exogenous) sick pay and the health state at which workers will absent. They assume that the firm has some prior over the distribution of health states and has some monitoring technology to identify 'shirkers' at some cost. The firm then chooses a wage and number of employment contracts to maximise expected profits. The conclusion of the paper is common to many simple efficiency wage models: if the cost of monitoring workers increases, the efficiency wage increases to ensure the appropriate no shirking constraint binds. Empirically this conclusion implies that the efficiency wage varies with firms' monitoring technology. The model also predicts that, for a given health state, workers are more likely to be absent: the higher the sick pay; the higher the unearned income; the higher the level of contracted hours; and the lower the wage. Consequently the empirical implications are rather similar to the standard labour supply approach.

2(iii) The Contract Approach

Coles and Treble (1993) adopt a contractual approach to absenteeism. Specifically, identical workers may be either: absent with cause; or choose to be absent without cause; or choose to be at work. The firm, however, only observes the absence-attendance choice of the worker. The problem for the firm is to choose some wage-sick pay contract to maximise profits subject to a zero profit condition and an incentive compatibility constraint (ICC). The ICC is specified to ensure that workers

who might be absent without good cause do in fact attend. The equilibrium is compared for both constant returns to scale firms and firms that employ an assembly line technology[18]. It is a standard moral hazard model that illustrates the trade off between incentives and insurance in firms with different technologies. The authors show that agency costs are higher (under certain circumstances) in an assembly line firm. Indeed this result is the main focus of the paper. The empirical predictions of the model are that assembly line firms are more likely to pay higher wages, lower sick pay and experience a lower rate of absence as a result. These predictions are again in line with those made by the supply side studies. Nonetheless Coles and Treble (op. cit., 1993) do provide a useful framework in which absenteeism is the outcome of the interaction between firm and worker choice. It is important to note, however, that only healthy workers have the discretion to choose to attend or not. The possibility that sick workers may choose to attend work in order to receive a higher remuneration is ruled out by assumption. In other words the possibility of presenteeism is ignored. It is therefore hard to see why firms should offer a strictly positive sick pay[19]. Furthermore, the authors impose a zero profit condition which, together with the incentive compatibility constraint, tie down the level of the wage and sick pay. The zero profit condition also implies that the firm is indifferent to contracts that yield high or low absence rates.

2(iv) Presenteeism and Sick Pay

Our paper makes the reasonable assumption that productivity is some function of health state. This requires a richer sorting than in Coles and Treble (op. cit., 1993). In particular, incentive compatibility considerations require greater complexity in specification. Unwarranted absenteeism must still be prevented by respecting a suitable incentive compatibility constraint which will deter 'healthy' workers from

absenting themselves. However, unlike all previous work, we also examine the other side of this coin, viz. unwarranted "presenteeism". Since productivity is some function of health state, 'sick' workers may need to be discouraged from attempting to attend. The implication for the determination of wages, (extra-statutory) sick pay, absence and empirical studies are analysed.

Section 3: Preventing Absenteeism and Presenteeism

We seek to generalise the contract approach by examining the implications of sickness, both on the behaviour of firms and of workers. Sickness is seen as a random phenomenon and a traditional principal-agent model is developed in which the agent can choose to attend work or choose to be absent, ostensibly on grounds of sickness.

The precise nature of the game we wish to analyse is as follows. The principal (firm) wishes to run a one-time project by employing an agent (worker) to produce some output. Output is assumed to depend on both the attendance, and health, θ, of the agent. There are only two possible states of health, and these health states are exogenously determined. These health states are denoted θ_H if 'ill' and θ_L if 'healthy', where the subscripts L and H refer to low and high θ, respectively.

The essential characteristic of the production process is assumed to be that attendance by unhealthy workers lowers output and thus firms may wish unhealthy workers to stay at home. There may be a number of reasons for this. For example, *Labour Research* (August 1998) reports that, 'the lowest possible absence rates are not necessarily the best outcome for a firm or organisation, as this may result in presenteeism. This means 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). Steers and Rhodes (1978)

reinforce this idea by noting, '…some absenteeism may in fact be healthy for organisations in that such behaviour can allow for temporary escape from stressful situations…' and that, 'rigid attempts to ensure perfect attendance may lead to unintended and detrimental consequences on the job', p. 34. Thus the notion that some firms may wish to separate its workforce in this fashion has already been noted, though perhaps not fully analysed.

We capture this idea in a very simple way by assuming that there is some critical value of sickness, $\overline{\theta}$, below which firms wish agents to attend and above which firms wish agents to stay at home. We further assume that the parameter values are such that $\theta_L < \overline{\theta} < \theta_H$. Thus, it is in the interest of the principal to ensure a separation of the type illustrated in the matrix below.

	θ_L	θ_H
Attend	Yes	No
Absent	No	Yes

Table 1: Labour allocation.

After the contract has been accepted, information regarding the agent's health is revealed to the agent only: he is assigned θ_H or θ_L. The distribution of θ is known. In particular, the probability of the agent being assigned θ_L is π and the probability of being assigned θ_H is $(1-\pi)$.

Given this information, the agent can choose to go to work and supply an exogenous amount of effort, e, or stay at home and absent himself on grounds of sickness and

supply, obviously, zero effort[20]. This information can be illustrated using the timeline below.
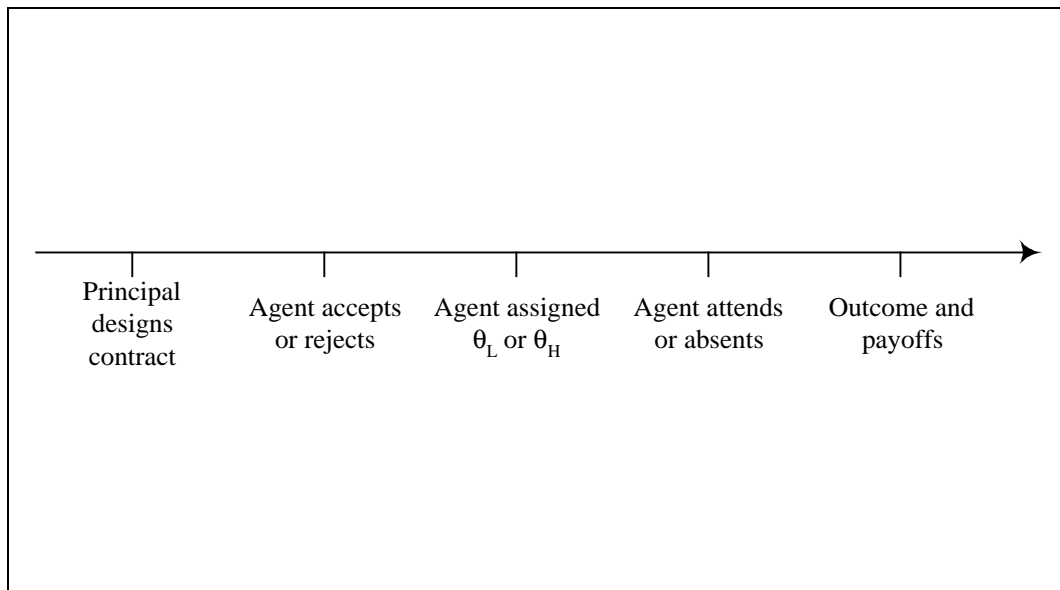

Figure 3: The structure of the problem.

Under the assumption that the principal wishes to separate the two types, the principal pays some wage, w, to those who attend and some sick pay, s, to those who absent themselves to maximise her utility.

Formally, when all healthy workers attend and all ill workers absent, the principal has an expected utility function of the form

$$V^P = \pi[x(e, \theta_L) - w] - (1 - \pi)s. \tag{1}$$

In (1) $\pi[x(e, \theta_L) - w]$ represents the profit to the principal of a healthy worker who attends with probability π, supplying exogenous effort, e^{21}, to produce output, x. While the second term in (1) represents the sickness payment to an absent worker, s, who is ill with probability $(1 - \pi)$.

An agent of type θ_i has utility

$$V^A = \begin{cases} u(w) - v(e, \theta_i) & \text{if the agent attends} \\ u(s) - v(0, \theta_i) & \text{if the agent is absent} \end{cases} \tag{2}$$

We assume the agent's utility is separable in reward and cost ($u(\cdot)$ and $v(\cdot)$)[22] both measured in terms of utility and that the following assumptions hold.

$$u'(\cdot) > 0; u''(\cdot) < 0, \tag{3}$$

$$v(\cdot, \theta_H) > v(\cdot, \theta_L), \tag{4}$$

$$v(e, \theta_i) > v(0, \theta_i), \tag{5}$$

$$v(e, \theta_H) - v(0, \theta_H) > v(e, \theta_L) - v(0, \theta_L). \tag{6}$$

Assumption (4) states that for any given level of effort, the lower the agent's health, the greater his disutility.

Assumption (5) states that, for a given health value, the disutility of going to work is strictly greater than the disutility of staying at home.

Assumption (6) makes the quite reasonable assumption that the additional disutility of going to work is higher for sick workers than for healthy ones. This is the discrete analogue of the Spence-Mirrlees condition and implies that those agents with higher values of θ have to be paid more for attendance than those agents with lower values

of θ. Coupled with the lower productivity of unhealthy workers, this provides both the opportunity and the incentive to separate the two types.

3(i) Full Information and Complete Contracting

We start by examining a benchmark case in which all information relevant to the contract, in particular health state, is observable and verifiable ex-post, so that complete contracts can be made contingent on this. The contract is accepted prior to the agent's health being known and we assume the agent is bound by the terms of the contract once his health state is revealed: the agent can't renege. In effect this ensures that a contract based on attendance will be equivalent to one based on health state. The principal has to choose two wages, (w,s) to maximise her utility subject to the constraint that the agent accepts the contract, i.e. the participation constraint (PC). The wage w is paid to attendees, whilst s is paid to absentees.

Since the contract is signed before health state is known, and the no reneging/health state verifiable assumptions mean that workers must behave in accordance with their true health state, the only constraint on the firms wage policy is the participation constraint. This is of the form 'expected utility from accepting the contract' exceeds or equals 'expected utility from rejecting the contract'. The expected utility from acceptance is:

$$\pi[u(w) - v(e, \theta_L)] + (1 - \pi)[u(s) - v(0, \theta_H)] \tag{7}$$

and if b is the income received from rejection of the contract, then the expected utility from rejection is :

$$\pi[u(b) - v(0, \theta_L)] + (1 - \pi)[u(b) - v(0, \theta_H)], \tag{8}$$

which can be interpreted as an expected reservation utility. The expected utility participation constraint can then be written,

$$\pi[u(w) - v(e, \theta_L) + v(0, \theta_L)] + (1 - \pi)u(s) \geq u(b). \tag{9}$$

Thus in this benchmark case the principal wishes to solve the following problem,

$$\underset{w,s}{\text{Max}} \quad \pi[x(e, \theta_L) - w] - (1 - \pi)s \tag{10}$$
$$\text{s.t.} \quad \pi[u(w) - v(e, \theta_L) + v(0, \theta_L)] + (1 - \pi)u(s) \geq u(b).$$

Forming the associated Lagrangian we have

$$\underset{w,s,\lambda}{\text{Max}} L = \pi[x(e, \theta) - w] - (1 - \pi)s \tag{11}$$
$$+ \lambda\{\pi[u(w) - v(e, \theta_L) + v(0, \theta_L)] + (1 - \pi)u(s) - u(b)\}.$$

It is clear from (1) that the PC will always hold with equality because the principal's utility is decreasing in w and s. There are four possible configurations of w and s- each can be either positive or zero. It can be shown that $w = 0$, $s > 0$ leads to a contradiction while $w = s = 0$ obviously violates the PC. Hence $w > 0$ and we are thus left with only two cases to consider, $s > 0$ and $s = 0$. The Kuhn-Tucker conditions are:

$$L_w: \quad -\pi + \lambda\pi u'(w) = 0 \tag{12}$$

$$L_s: \ -(1-\pi) + \lambda(1-\pi)u'(s) \begin{cases} \leq 0 \\ = 0, \ s > 0 \end{cases} \tag{13}$$

$$L_\lambda: \ \pi[u(w) - v(e,\theta_L) + v(0,\theta_L)] + (1-\pi)u(s) = u(b). \tag{14}$$

Consider first, the case of $s = 0$. Combining (12) and (13) we have

$$\frac{1}{u'(w)} \leq \frac{1}{u'(0)}. \tag{15}$$

which implies non-diminishing marginal utility of income since $w > 0$. This violates strict concavity of $u(\cdot)$. Thus $s = 0$ leads to a contradiction. Hence, s is strictly positive and the solution to the above problem can therefore be written

$$L_w: \ -\pi + \lambda\pi u'(w) = 0 \tag{16}$$

$$L_s: \ -(1-\pi) + \lambda(1-\pi)u'(s) = 0 \tag{17}$$

$$L_\lambda: \ \pi[u(w) - v(e,\theta_L) + v(0,\theta_L)] + (1-\pi)u(s) = u(b). \tag{18}$$

(16) and (17) imply

$$\lambda = \frac{1}{u'(w)} = \frac{1}{u'(s)} \Rightarrow w = s \tag{19}$$

and substitution of (19) in (18) implies the optimal values $\{\hat{w}, \hat{s}\}$ satisfy

$$u(\hat{s}) = u(\hat{w}) = u(b) + \pi[v(e, \theta_L) - v(0, \theta_L)]. \tag{20}$$

From (20), the optimal payments are increasing in b (the alternative value of employment), π (the probability of being healthy) and e (the minimum effort requirement at work).

Coles and Treble (op. cit., 1993) characterise this contract as the 'high absence' contract. This contract, in which $w = s$, ensures Pareto efficiency since risk averse workers prefer state independent wage contracts[23]. In this complete contractibility/no reneging case, the standard results of optimal risk sharing apply: a risk neutral principal fully insures a risk averse agent by making payments independent of the agent's state of health. In addition, ex ante, the agent's expected utility is equal to his reservation utility since the participation constraint binds. See figure 4 below.
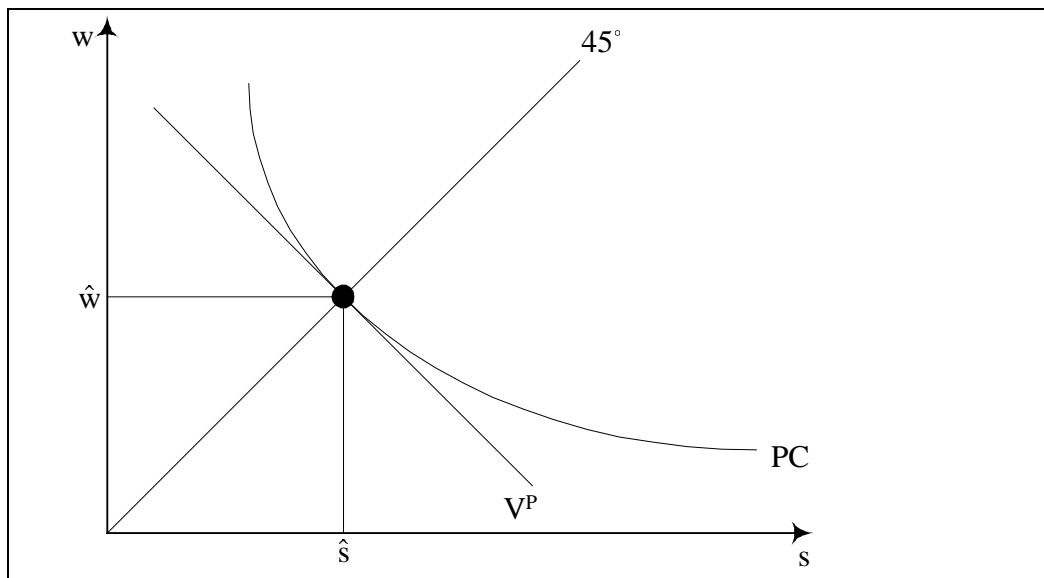

Figure 4: The Benchmark case.

If healthy types observe the terms of the contract and present themselves for work their ex post utility is

$$u(\theta_L) = u(\hat{w}) - v(e, \theta_L) , \tag{21}$$

But his ex post utility if he could renege and absent himself is

$$u^r(\theta_L) = u(\hat{s}) - v(0, \theta_L) , \tag{22}$$

Since $\hat{w} = \hat{s}$ but $v(e, \theta_L) > v(0, \theta_L)$, healthy types have a strong incentive to renege and be absent. By contrast, the opposite is true for ill workers and they would actually lower their ex post payoff by misrepresenting themselves. This poses serious doubts about the practical usefulness of the full insurance wage policy in which w and s are equal. Except in special and extreme circumstances health state is not verifiable and hence the possibility of reneging must be seriously considered. If reneging is possible and the full insurance wage policy is offered both sick and healthy workers will stay at home. The result is extreme absenteeism and collapsed output. It is therefore of some interest to study the characteristics of a contract which separates the types when health states are not verifiable and reneging is possible. It is precisely this sort of 'cheating' by the healthy types that one would wish to control.

3(ii) Incomplete Contracting

Suppose now that that health state is unverifiable. It is only after the contract is signed that the agent becomes aware of his type, while the principal knows only the distribution of the two types, i.e. π is common knowledge.

As before, the principal can offer a menu of contracts {w,s} where w is the contract intended for the θ_L types and s is intended for the θ_H types. The principal only observes (and is able to verify to a third party) the agent's attendance decision which is contingent upon his θ. The principal thus makes the two contracts contingent upon attendance. This is in contrast to the symmetric information case where payment can be made contingent upon the agent's health state.

Because the principal is unable to observe the agent's true type, she must design a contract in such a way as to induce the agent to truthfully reveal his type and select the contracts designed specifically for that type. In other words, the contracts must be self-selective: each agent obtains greater utility from truthfully revealing their type than from cheating. The principal can achieve this self-selection by employing direct, or revealing, mechanisms[24].

The self-selection or incentive compatibility constraints can be written as follows. For type θ_L

$$u(w) - v(e, \theta_L) \geq u(s) - v(0, \theta_L), \tag{23}$$

such that if $\theta = \theta_L$ the agent prefers to attend and reveal his type to the principal by his choice of action. Similarly, for type θ_H the appropriate incentive compatibility constraint is

$$u(s) - v(0, \theta_H) \geq u(w) - v(e, \theta_H). \tag{24}$$

The principal's problem is then to maximise (1) subject to (9), (23) and (24),

The maximisation programme yields several combinations of the choice variables from which to search for possible solutions to the problem. Fortunately one can significantly reduce the possible configurations by a closer analysis of the assumptions and the constraints.

From assumption (5) and (23) we have

$$u(w) - u(s) \geq v(e, \theta_L) - v(0, \theta_L) > 0. \tag{25}$$

From which we can deduce that $w > 0$ and $w > s$.

Furthermore, combining (23) and (24) yields

$$v(e, \theta_H) - v(0, \theta_H) \geq u(w) - u(s) \geq v(e, \theta_L) - v(0, \theta_L). \tag{26}$$

Using the strict inequality in (6), the Spence-Mirrlees condition, at least one of the inequalities in (26) must be strict. Moreover, both incentive compatibility constraints cannot be strict inequalities: otherwise a reduction in w or s would satisfy both constraints while increasing the utility of the principal.

The solution to this programme can be found using the Kuhn-Tucker conditions. But a more intuitive approach is through the use of the following graphical analysis as shown in figure 5.

The slope of the principal's objective function in (s, w) space is

$$\left. \frac{dw}{ds} \right|_{V^P} = \frac{-(1-\pi)}{\pi}. \tag{27}$$

While the slope of the agent's expected participation constraint is

$$\left.\frac{dw}{ds}\right|_{PC} = \frac{-(1-\pi)u'(s)}{\pi u'(w)} . \tag{28}$$

Note that the slopes of the above curves are the same for $w = s$, or along the $45°$ line. In addition, we know from (27) and (28) that for $w > s$, the slope of the principal's objective function is less than the slope of the expected participation constraint, by concavity of $u(\cdot)$. Hence for any incentive compatibility constraint that has an intercept with the vertical axis lower than the participation constraint, we should have a solution in which $s > 0$ [25]. Furthermore it can be shown that the ICC for the 'sick' types- which we call IC_H- does not bind[26].

So, in order for the principal to induce the agent to select the contract designed specifically for their health value, she introduces a distortion from the full insurance solution by increasing the spread between w and s. The agent now bears some income risk.

Formally, this solution satisfies the following Kuhn-Tucker conditions for $\{w > 0; \ s > 0; \ \lambda > 0; \ \mu > 0; \ \eta = 0\}$ [27]

$$L_w: \ u'(w)[\lambda\pi + \mu] - \pi = 0 \tag{29}$$

$$L_s: -(1-\pi) + u'(s)[\lambda(1-\pi) - \mu] = 0 \tag{30}$$

$$L_\lambda: \ \pi[u(w) - v(e, \theta_L) + v(0, \theta_L)] + (1-\pi)u(s) - u(b) = 0 \tag{31}$$

$$L_\mu: u(w) - v(e,\theta_L) - u(s) + v(0,\theta_L) = 0 \tag{32}$$

$$L_\eta: u(s) - v(0,\theta_H) - u(w) + v(e,\theta_H) > 0. \tag{33}$$

Where the optimal solutions $\{\tilde{w}, \tilde{s}\}$ satisfy

$$u(\tilde{w}) = u(b) + v(e,\theta_L) - v(0,\theta_L) \tag{34}$$

and

$$u(\tilde{s}) = u(b). \tag{35}$$

From which we have

$$\begin{aligned} \tilde{w} &> \hat{w} \\ \tilde{s} &< \hat{s} \end{aligned}. \tag{36}$$

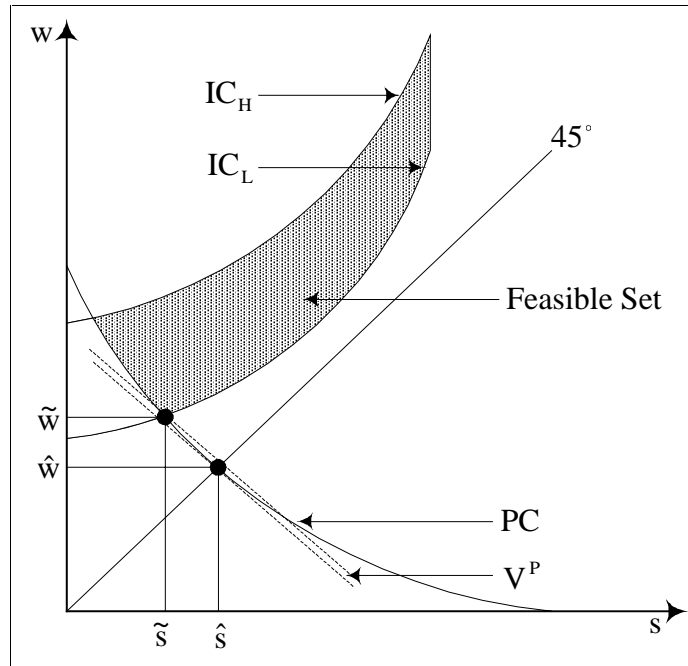Figure 5 below illustrates the two contracts.

Figure 5: A comparison of the different contracts.

Neither type now has an incentive to masquerade as another. Healthy workers now have no incentive to cheat since their ICC binds whereas sick workers would actually be worse off by misrepresentation as before. The contract has achieved a complete sorting of types but at the cost of distorting full insurance.

Section 4: Empirical Implications

The results of the analysis rest on the following assumptions. Firstly, that ill agents are sufficiently less productive than healthy individuals so that the firm wishes them to stay at home. Secondly, that illness is non-contractible. Under what circumstances are these assumptions likely to be empirically valid? If the firm were unconcerned about ill workers attending, it has no obvious reason to offer a strictly positive sickness pay[28]. This is not observed empirically. Furthermore, sickness absence in the UK is contractible to some degree. Specifically, sickness absence for duration over three days has to be certified (verified) by a qualified medical practitioner. The three

days for which illness is self-certified however yields a significant amount of non-contractible absence. Labour Market Trends reports that in winter[29] 1999/2000 49% of respondents were unable to work for 3 days or less as a result of sickness or injury. Therefore the agency approach adopted in this paper has some validity.

It is important to recognise that there is a perfect sorting of types in this model. Conventional comparative statics exercises therefore affect only the optimal contract $\{\tilde{w}, \tilde{s}\}$ and not the absence rate which, at the optimum, is determined by the probability of illness, $1 - \pi$. It is clear from (34) and (35) that the optimal contract is given by:

$$\tilde{w} = \tilde{w}[\overset{+}{b}, \overset{+}{v(e,\theta_L)} - v(0,\theta_L)] \atop \tilde{s} = \tilde{s}[\overset{+}{b}] \tag{37}$$

A more informative approach to adopt would be to analyse the model in dis-equilibrium regimes. Consider figure 6 below. The feasible set of contracts is the shaded area bounded by the lower envelope of IC_H and the upper envelopes of IC_L and the participation constraint. One of the firm's isoprofit curves, V^P, has also been drawn. In addition, the wage at which IC_H just binds when $s = \tilde{s}$ is denoted by $\overline{w}$.
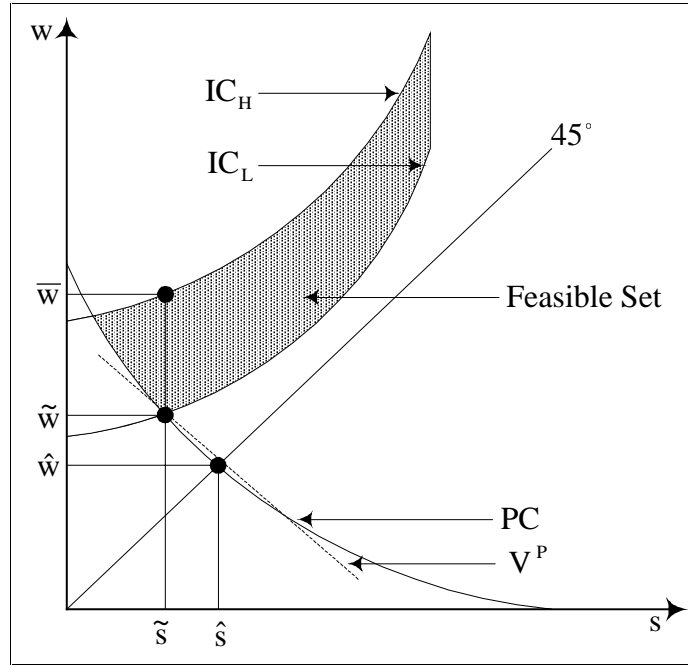
Figure 6: Disequilibrium regimes

In table 2 we fix s at $\tilde{s}$ and consider the potential 'mistakes' the firm may make in its choice of w. Four possible cases are analysed.

	$s = \tilde{s}$ and	PC satisfied	$\theta = \theta_H$	$\theta = \theta_L$	Absence rate	Rent
(1)	$w < \tilde{w}$	No				
(2)	$w = \tilde{w}$	Yes	Absent	Attend	$1 - \pi$	No
(3)	$\tilde{w} \le w \le \overline{w}$	Yes	Absent	Attend	$1 - \pi$	Yes
(4)	$w > \overline{w}$	Yes	Attend	Attend	0	Yes

Table 2: Potential 'mistakes' when $s = \tilde{s}$

For example, suppose the firm correctly selects the optimal level of sick pay $s = \tilde{s}$ but errs in its choice of w. Specifically, the firm chooses some wage $\tilde{w} \le w \le \overline{w}$. This corresponds to row three in table 2 above. While obviously sub-optimal, this point is in the feasible set. If the agent is ill, $\theta = \theta_H$, he absents. If the agent is healthy,

$\theta = \theta_L$, he attends work. The sorting of types is still perfect and the absence rate is optimal. The only difference is that agents are now paid a rent: the PC doesn't bind.

Now consider the case in which the firm chooses the correct wage, $\tilde{w}$, but errs in its choice of sick pay, s.

	$\mathbf{w} = \tilde{\mathbf{w}}$ and	PC satisfied	$\mathbf{\theta} = \mathbf{\theta}_\mathbf{H}$	$\mathbf{\theta} = \mathbf{\theta}_\mathbf{L}$	Absence rate	Rent
(1)	$s < \tilde{s}$	No				
(2)	$s = \tilde{s}$	Yes	Absent	Attend	$1 - \pi$	No
(3)	$s > \tilde{s}$	Yes	Absent	Absent	1	Yes

Table 3: Potential 'mistakes' when $\mathbf{w} = \tilde{\mathbf{w}}$

The implication of a mistake in the choice of s, given $w = \tilde{w}$, are particularly stark. If the firm correctly sets $s = \tilde{s}$, some production takes place. Otherwise output is zero. In welfare terms the case in which $s < \tilde{s}$ is preferable to the case in which $s > \tilde{s}$, since in the former there are no employment costs because no agent signs the contract. However in the latter case, the worker signs the employment contract; the firm pays sick pay to all employees; but no production takes place.

Thus the empirical implication of estimating a linear regression equation of the form Absence Rate $= \alpha + \beta w + Z\delta$ (e.g. Barmby and Stephan (2000), Brown et al. (1999)) are far from clear. Indeed, the significance of β depends crucially upon the regime in which the contract is located. For example, for $s = \tilde{s}$ there is a range of w, $\tilde{w} \leq w \leq \overline{w}$, within which absence is invariant to changes in the wage.

These implications have been drawn from the model assuming a representative agent. A simple way to introduce a degree of heterogeneity into the model would be to suppose that agents differ in $v(e, \theta_H) - v(0, \theta_H)$ only. This is a particularly simple case to analyse as it only affects IC_H.

Let $v(e, \theta_H) - v(0, \theta_H) = \alpha$, be distributed on $[\alpha^-, \alpha^+]$ with cdf $F(\cdot)$. For any remuneration strategy of the firm (w, s) it follows from IC_H (equation 24) that of the sick workers only those with $\alpha \geq u(w) - u(s)$ will properly absent themselves whilst those with $\alpha \leq u(w) - u(s)$ will improperly attend work. Hence the absence rate will be $1 - F[u(w) - u(s)]$ whilst the rate of presenteeism will be $F[u(w) - u(s)]$. These workers represent a genuine cost to the firm who would prefer them to stay at home. Note that the impact of wages on absence is more complex than simple linear regressions of the type frequently estimated in the literature might suggest. In particular, both absence and presenteeism depend on the interaction between w and s through $u(w) - u(s)$.

Section 5: Conclusions

In this paper we have extended the analysis of absence and sick pay by considering the obvious impact on productivity of work attendance by sick workers. This is the phenomenon sometimes called presenteeism. The resulting contract embodies a richer complex of incentive compatibility considerations. We are able to show that the optimal contract (which distorts full insurance) involves firms offering: (1) a wage that is strictly higher than the sickness pay offer, and (2) a sickness pay that is strictly positive.

Our results suggest that empirical absence functions conventionally estimated in the literature may be seriously mis-specified. Our results also shed a slightly different light on the concerns regarding absence cited in the introduction. One reaction to reducing absence may be to lower sickness pay. Our results suggest that such a policy may well result in greater presenteeism. More broadly, any attempt at reducing the potential productivity loss from absence has to be offset against the potential

productivity loss from presenteeism. Since the significance and extent of presenteeism is virtually unknown more empirical work in this area is clearly called for. Finally, a richer specification of the production process in which the firm's cut-off point $\bar{\theta}$ is determined from more fundamental considerations and the extension of the analysis to a multi-period framework would be desirable.

References

Allen, S. G., 1981, "An empirical model of work attendance", *The Review of Economics and Statistics,* **63,** 77-87.

Allen, W. D., 1996, "Family illness and temporary work absence", *Applied Economics,* **28,** 1177-1180.

Barmby, T. and Stephan, G., 2000, "Worker absenteeism: why firm size may matter", *The Manchester School*, **68**, 568-577.

Barmby, T. A., Ercolani, M. O. and Treble, J. G., 1999, "Using the Labour Force Surveys to construct an absence rate series for Britain, 1984-1997 (revised)", *Institute for Labour Research Working Paper*, No. 27.

Barmby, T. A., Orme, C. D. and Treble, J. G., 1991, "Worker absenteeism: an analysis using microdata", *Economic Journal,* **101,** 214-229.

Barmby, T. A. and Treble, J. G., 1991, "Absenteeism in a medium-sized manufacturing plant", *Applied Economics,* **23,** 161-166.

Barmby, T., Sessions, J. and Treble, J., 1994, "Absenteeism, efficiency wages and shirking', *Scandinavian Journal of Economics,* **96,** 561-566.

Brown, S., Fakhfakh, F., and Sessions, J., 1999, "Absenteeism and employee sharing: an empirical model based on French panel data", *Industrial and Labor Relations Review,* **52,** 234-251.

Brown, S. and Sessions, J. G., 1996, "The economics of absence: theory and evidence", *Journal of Economic Surveys,* **10,** 23-53.

Brown, S., 1994, "Dynamic implications of absence behaviour", *Applied Economics,* **26,** 1163-1175.

Coles, M. G. and Treble, J. G., 1993, "The price of worker reliability', *Economics Letters,* **44,** 149-155.

Confederation of British Industry, 1997, "Managing absence- in sickness and in health", *CBI Publications Unit,* London, .

Deardorff, A. V. and Stafford, F. P., 1976, "Compensation of co-operating factors", *Econometrica,* **44,** 671-684.

Drago, R. and Wooden, M., 1992, "The determinants of labor absence: economic factors and workgroup norms across countries", *Industrial and Labor Relations Review,* **45,** 764-778.

Dunn, L. F. and Youngblood, S. A., 1986, "Absenteeism as a mechanism for approaching an optimal labor market equilibrium: an empirical study", *The Review of Economics and Statistics,* **68,** 668-674.

Kahana, N. and Weiss, A., 1992, "Absenteeism: a comparison of incentives in alternative organisations", *Journal of Comparative Economics,* **14,** 573-595.

Killingsworth, M., 1983, "Labor supply", Cambridge, Cambridge University Press.

Labour Market Trends, May 2000, 190.

Labour Research, August 1998, "Noticeable by their absence", 21-22.

Lazear, E. P., 1981, "Agency, earnings profiles, productivity, and hours restrictions", *American Economic Review,* **71,** 606-620.

Lazear, E. P., 1979, "Why is there mandatory retirement?', *Journal of Political Economy,* **87,** 1261-1284.

Mas-Colell, A., Whinston, M. D. and Green, J. R., 1995, "Microeconomic theory", Oxford, Oxford University Press.

Myerson, R. B., 1979, "Incentive compatibility and the bargaining problem", *Econometrica,* **47,** 61-73.

Steers, R. M. and Rhodes, S. R., 1978, "Major influences on employee attendance: a process model", *Journal of Applied Psychology,* **63,** 391-407.

Stewart, M. B. and Swaffield, J. K., 1996, "Constraints on the desired hours of British men', Warwick Economic Research Papers, Number 468.

The Economist, 10th October 1998, "Sick of work: cracking down on malingerers".

The Times, 23rd July 1998, "Public sector sickness crackdown", 1.

Viscusi, W. K. and Evans, W. N., 1990, "Utility functions that depend on health status: Estimates and economic implications", *The American Economic Review,* **80,** 353-374.

Weiss, A., 1985, "Absenteeism and wages", *Economics Letters,* **19,** 277-279.

[1] CBI (1997) exhibit 12, p. 13.

[2] Stewart and Swaffield (1996) provide estimates of the magnitude of this difference.

[3] There are no discontinuities in the budget set facing the individual.

[4] With the exception of corner solutions to the problem.

[5] Deardoff and Stafford (1976) and Drago and Wooden (1992) cite a number of reasons that firms wish to co-ordinate production to some degree. Lazear (1981) shows that hours restrictions are one component of an optimal employment contract when age-earning profiles are upward sloping as in Lazear (1979).

[6] Stewart and Swaffield (1996) estimate that around one third of men in their sample work longer hours than they would choose at the prevailing wage. They estimate that men would choose to work 4.3 hours less per week, on average.

[7] The variables w and z denote the hourly wage and unearned income, respectively; L and C denote leisure hours and consumption, respectively; and h^c and T denote an hours and total time constraint, respectively.

[8] If $MRS^c < w/p$, the individual has an incentive to supply more labour by taking an extra job or moonlighting.

[9] Note the difference between the two notions referred to here. On the one hand the difference between an individual's marginal rate of substitution and the wage rate determines the agent's incentive to absent. On the other hand, absence is a decision or an action rather then merely a temptation.

[10] Killingsworth (1983) discusses whether a point such as A can nevertheless be described as an equilibrium in the sense that there is no tendency for either firm or employee to alter their decisions, despite the solution being suboptimal for the worker. Or, in terms of Killingsworth (1983), the firm and worker choose not to exploit any potential gains from trade.

[11] Or, more precisely, absence pay, since sickness or ill health is private information.

[12] Presumably, sick pay is targeted only at 'sick' individuals.

[13] From the work of Viscusi and Evans (1990)

[14] This is the intuition underlying Allen's (1996) analysis of absence. He contends that workers with families bear a larger psychic cost of work and thus absent more frequently and for longer durations in the event of family illness. His empirical results support this idea that the MRS is greater for those with families. Barmby et al. (1991) also find that gender and marital status are important determinants of absence.

[15] This property is extremely important in the agency model that follows. Where the health of the agent can be represented as a continuous variable, this property of the utility function ensures that indifference curves cross only once for different health states. This single crossing property gives rise to the possibility that the firm (or a principal) can separate agents on the basis of their health. This is termed the Spence-Mirrlees, or sorting, condition that allows agents to be sorted according to their type.

[16] Assuming leisure is a normal good.

[17] This strategic interaction has been recognised in the literature. Dunn and Youngblood (1986) report that, "workers were using illness absence for non-illness purposes because it provided them with paid time off that was easily legitimised as illness."

[18] Coles and Treble (1993) essentially generalise the approach of Weiss (1985).

[19] Indeed Coles and Treble (op. cit., 1993) do not demonstrate the positivity of the sick pay.

[20] i.e. these decisions represent the corner solutions to the (daily) labour supply problem considered in the introduction.

[21] The level of effort here is arbitrary and can be thought of as the effort required to go to work. Thus it simply represents attendance.

[22] Given this separation of the utility unction we are unable to capture decreasing marginal utility of money, as in Viscusi and Evans (1990).

[23] This is similar to the 'high absence' contract of Coles and Treble (op. cit., 1993)

[24] The Revelation principle asserts that the principal may restrict her search for the optimal contract to those that provide each agent with the incentive to truthfully reveal his characteristics. See Mas-Colell et al. (1995) or Myerson (1979) for a formal description.

[25] For $u(0) = 0$ and (5), the intercept of IC_L is lower than the intercept of PC.

[26] If IC_H binds, we arrive at the contradiction that $u'(w)$ exceeds $u'(s)$.

[27] The Kuhn-Tucker conditions are necessary and sufficient if IC_L (the binding constraint) is convex. This is true for $u'''(\cdot) \le 0$.

[28] In fact the model of Coles and Treble (1993) implies this. For any {w, s} contract it can be shown that the optimal choice of sick pay is 0.

[29] Barmby et al. (1999) show that absence is higher during the winter months than on average.