Theoretical approaches to managing sickness absenteeism

Chatterji, Monojit; Tilley, Colin

Publication date:
2003

University
of
Dundee

Dundee Discussion Papers in Economics

Theoretical Approaches to Managing Sickness Absenteeism

Monojit Chatterji
and
Colin Tilley

Department of
Economic Studies,
University of Dundee,
Dundee.
DD1 4HN

Working Paper
No. 143
May 2003
ISSN:1473-236X

Theoretical approaches to managing sickness absenteeism

Monojit Chatterji
Department of Economic Studies
University of Dundee

Colin Tilley*
DHSRU
University of Dundee

Abstract

The cost of absenteeism in the UK has been estimated to be £12 billion per annum. If productivity is some function of the health state of the worker firms may prefer some sickness absence to universal attendance. However, when the health state of the worker cannot be verified the firm must structure its employment contract in order to align the workers' incentives with its own. The nature of the optimal contract under these circumstances has recently been analysed by Chatterji & Tilley (2002) and Skåtun (2003) who generate rather different theoretical results and empirical implications. In this paper we synthesise these two approaches and reconcile their results.

JEL Classification: J22, J33, J41

Keywords: Absenteeism, incomplete information, wages.

*c.j.tilley@dundee.ac.uk. *Tel.:* +44 (0) 1382 425753. *Fax:* +44 (0) 1382 226550

1 Introduction

The latest CBI survey on absenteeism and labour turnover (CBI (2002)) estimates that absence from work in the UK cost business approximately £12 billion (1% of GDP) in 2001. This considerable output loss due to absenteeism has been the main focus of much previous literature, most of which has been empirical in nature. In analysing the causes and consequences of absenteeism this literature has viewed absence as a form of rational shirking and hence undesirable from the perspective of both firms and society.

More recently a new perspective has been emerging which recognises explicitly that there is a link between productivity and health state of the worker. This productivity-health link immediately suggests that not all absenteeism is universally undesirable. Specifically, this new approach posits that firms may find it in their interest to discourage unhealthy, low productivity workers from attending because such workers may well have a low productivity. Thus, just as 'healthy' workers absenting themselves is damaging, so might the situation in which 'unhealthy' workers turn up for work.

A natural question which arises in this framework is what policies the firm may engage in order to control both the quantity of absenteeism and its flip side – unhealthy workers turning up to work. In essence firms have to design employment contracts in such a way that healthy workers are encouraged to turn up for work whilst sick workers are encouraged to stay at home. If health state is not 'contractible' – verifiable to a third party – firms who wish to sort workers according to their health state have to design an employment contract that is incentive compatible otherwise healthy workers may falsely claim to be ill and ill workers may falsely claim to be healthy.

The nature of the optimal employment contract under these circumstances is not immediately obvious. The issue has recently been independently analysed by

Chatterji & Tilley (2002) and Skåtun (2003) who generate rather different theoretical results and empirical implications. Chatterji & Tilley (2002) argue that the optimal employment contract involves the firm setting two wage rates only – one for those who attend work and one for those who stay at home. By contrast Skåtun (2003) suggests that it is best to offer three wage rates. One for healthy workers who attend, one for sick workers instructed to attend work, and one for all workers who stay at home. Furthermore, in Chatterji & Tilley (2002) full sorting occurs, whereas in Skåtun (2003) some sick workers do present themselves for work so that imperfect sorting occurs. This short paper synthesises the two approaches and shows that the differences in conclusions arise from different assumptions about (i) the information that is verifiable ex ante and (ii) the variables upon which payment is conditioned. In particular, we derive the optimal employment contract for a separating equilibrium using two different information sets.

2 Absenteeism in the UK

The responsibilities of firms and workers when workers claim sickness absence is determined in part by the Department of Social Security sickness absence regulations and in part by the policies of the firm. The Department of Social Security only requires a doctor to issue a statement to certify illness after an individual has been absent for a week. In other cases, a worker may simply claim to be ill, i.e. to self certify. If the individual is claiming sickness absence for at least 4 days in a row then they can claim statutory sick pay (SSP) which is payable for up to 28 weeks in one 'spell' of absence. After the 28 week period is over, the individual is transferred to Incapacity Benefit. Sickness absence periods of at least 4 days in a row with 8 weeks or less between them are counted as one spell of absence for SSP purposes. SSP is not payable in the first 3 days (the waiting days) of any spell.

Barmby, Orme & Treble (1991) finds that the majority of sickness absence is in spells of five days or less. Furthermore, using data from the Labour Force Survey, LMT (2003) show that of those workers who were absent during a reference week, 40% of workers claimed absence for a period of only one day and approximately 75% claimed absence for 4 days or less. Both these suggest strongly that much absenteeism is on the basis of self certification of illness. Hence it would appear that firms are faced with self certified illness based absence and therefore need to consider wage and sick pay policies to align the incentives of workers with their own. Employers may well choose to provide their own company sick pay in addition to SSP. Indeed, Brown (1994) reports that in a survey of 1125 private and public sector employers over 70% offered company sick pay schemes as opposed to relying exclusively on the statutory sick pay scheme. Below we analyse these contractual arrangements using a formal model.

3 The model

For simplicity we synthesise Chatterji & Tilley (2002) and Skåtun (2003) within a static framework without loss of generality. A firm – the principal – is assumed to contract with an agent to run a one-time project. The agent may be either ill, θ_I, or healthy, θ_H, with associated probabilities ρ and $(1-\rho)$ respectively. Only after the contract between the parties has been signed is the agent's health state, θ, revealed to workers. These workers then decide whether to turn up to work or not depending on their health state, the monetary rewards, and the disutility of working given their health states. An agent with health state θ_j suffers disutility $v(e, \theta_j)$ if they voluntarily present themselves at work and supply exogenous effort e; they suffer disutility $v(0, \theta_j)$ if they are absent and consequently supply 0 effort.

However, in order to discourage shirking, the firm randomly selects a proportion, $(1-\phi)$, of absent workers and forces them to work. In terms of pay setting a crucial

question is whether or not the firm has knowledge of the ex post health state of any worker. The key assumption in Skåtun (2003) is

> 'that a firm can observe the productivity of a worker who turns up for work, thus making it impossible for a sick worker to successfully imitate a healthy worker at work.' (p. 391).

This immediately implies that the firm's employment contract can discriminate between healthy workers who voluntarily present for work, and those ill workers who, with probability $(1-\phi)$ will be press ganged into working. We denote these payments by w_H and w_I respectively. The payment to workers who are absent from work (the sickness pay) is denoted by w_S.

An agent who is healthy (H) and presents for work (P) receives utility $V_H^P = u(w_H) - v(e, \theta_H)$ while an agent who is healthy (H) and is absent from work (A) receives utility $V_H^A = u(w_S) - v(0, \theta_H)$. In contrast, an agent who is ill (I) and is instructed by the firm to present for work receives utility $V_I^P = u(w_I) - v(e, \theta_I)$ while an agent who is ill and absent receives utility $V_I^A = u(w_S) - v(0, \theta_I)$.

We assume that for any given level of effort the disutility when ill is greater than when healthy, $v(e, \theta_I) > v(e, \theta_H)$ and $v(0, \theta_I) > v(0, \theta_H)$; for a given health state the disutility of going to work is greater than the disutility of staying at home, $v(e, \theta_j) > v(0, \theta_j)$; and that the additional disutility of going to work when ill is greater than when healthy, $v(e, \theta_I) - v(0, \theta_I) > v(e, \theta_H) - v(0, \theta_H)$.[1]

Therefore the participation constraint of an agent is

$$(1-\rho)\left[u(w_H) - v(e, \theta_H)\right] + \rho(1-\phi)\left[u(w_I) - v(e, \theta_I)\right] + \rho\phi\left[u(w_S) - v(0, \theta_I)\right] \geq \bar{V} \tag{1}$$

[1] This is the discrete analogue of the Spence-Mirrlees condition and implies that θ_I agents have to be paid more for attendance than θ_H agents. Coupled with the lower productivity of unhealthy workers, this provides both the opportunity and the incentive to separate the two types of agent.

where $\bar{V}$ denotes the outside option of the agent.[2]

Denoting the productivity of a θ_j agent by $x(e, \theta_j)$ the expected value of the principal can be written

$$(1-\rho)\left[x(e,\theta_H) - w_H\right] + \rho(1-\phi)\left[x(e,\theta_I) - w_I\right] - \rho\phi w_S. \tag{2}$$

The nature of the resulting contract depends on whether full information or limited information is assumed.

3.1 Full information

Under full information the problem facing the principal is to choose w_H, w_I, and w_S to maximise expected profits (Equation 2) subject to the participation constraint (Equation 1).

If, ex ante, the principal is able to write a contract on the basis of both the health state and presence of the agent at work the standard optimal risk sharing results apply: a risk neutral principal will completely insure a risk averse agent against ex post fluctuations in utility. Therefore we have

$$u(w_H) - v(e,\theta_H) = u(w_I) - v(e,\theta_I) = u(w_S) - v(0,\theta_I). \tag{3}$$

Since $v(e,\theta_I) > v(e,\theta_H)$ Equation 3 implies

$$w_I > w_H. \tag{4}$$

[2] One of the innovations in Skåtun (2003) is that – because ill agents at work are more likely to infect others – the probability of illness, ρ, is an increasing function of the probability that an ill agent is forced to work. However, the firm knows this probability, $\rho(\phi)$, and internalises this infection 'externality'. Therefore, this is equivalent to assuming that from the perspective of workers the probability of being healthy is fixed.

Since $v(e, \theta_I) > v(0, \theta_I)$ Equation 3 implies

$$w_I > w_S. \tag{5}$$

If we further assume that $v(e, \theta_H) < v(0, \theta_I)$, then $w_S > w_H$.

Thus in the full information case the results of Skåtun (2003) regarding the relative magnitude of the three wages is confirmed. In effect, each worker is paid a wage that fully reflects their health state and work attendance status.

3.2 Incomplete information

In general, contracts cannot be made fully contingent on health status. The possibility that agents can try and bluff about their true health state in order to gain advantage must be taken into account by the firm. Typically this will impose a set of incentive compatibility constraints which are designed to prevent workers exploiting the informational advantage they have about their health status.

The two incentive compatibility constraints Skåtun (2003) imposes on the principal's problem are designed to ensure that healthy agents will always claim to be healthy rather than ill and to ensure that ill agents will always claim to be ill rather than healthy These are:

$$u(w_H) - v(e, \theta_H) \geq (1 - \phi)[u(w_I) - v(e, \theta_H)] + \phi[u(w_S) - v(0, \theta_H)]; \tag{6}$$

and

$$(1 - \phi)[u(w_I) - v(e, \theta_I)] + \phi[u(w_S) - v(0, \theta_I)] \geq u(w_H) - v(0, \theta_I). \tag{7}$$

Recall that the key assumption in Skåtun (2003) is

Assumption 1 'that a firm can observe the productivity of a worker who turns up for work, thus making it impossible for a sick worker to successfully imitate a healthy worker at work.' (p. 391).

Assumption 1 enables the principal to identify – and condition payment upon – the health state of workers who attend. To see this consider a healthy worker who claims to be ill. There is some probability, ϕ, that the agent's shirking will not be discovered. However, there is some probability, $(1-\phi)$, that the agent will be forced to attend in which case, by Assumption 1, they will be revealed to be healthy and should receive w_H.[3] Therefore the incentive compatibility constraint that ensures healthy agents attend (Inequality 6) becomes

$$u(w_H) - v(e, \theta_H) \geq u(w_S) - v(0, \theta_H). \tag{8}$$

Furthermore, since $v(e, \theta_j) > v(0, \theta_j)$, it follows that $w_H > w_S$ which is contrary to Skåtun (2003).

Moreover, if Inequality 8 is satisfied all healthy workers attend and payment to ill agents can be conditioned on health state. For example, an ill agent who is forced to attend with probability $(1-\phi)$ is identified through Assumption 1 and an ill agent who is absent is also identified because Inequality 8 ensures that a healthy agent will always attend. Therefore provided Inequality 8 is satisfied the second incentive compatibility constraint (Inequality 7) is redundant. The intuition for this result is that the principal always wants healthy workers to attend. Inequality 8 ensures this. In addition, the principal is only concerned that the right number of ill workers (if any) attend. Since ill health is verifiable the principal simply selects the number of ill agents required whether the ill agent is present or absent.

[3] Unless agents are fined for misreporting their health state.

In summary, if the health state of an ill agent is verifiable:

- only one incentive compatibility constraint is required
- the optimal contract consists of three wages that maximise the principal's objective function subject to the incentive compatibility constraint (8) and the participation constraint which both bind with equality.[4] However, the relative magnitude of these three wages do not accord with Skåtun (2003).

The assumption that healthy workers who turn up for work voluntarily and ill workers who are press ganged are paid different wages has little or no empirical support. If health state at work cannot be verified costlessly, then there is no possibility of there being separate wages for healthy and ill workers. The optimal contract in a separating equilibrium must consist of only *two* (rather than 3) wages. This is because the principal can only structure a contract on the basis of *verifiable* variables. In this case the principal can only verify presence at or absence from work.

In addition, in order to ensure a complete sorting of agents according to their health state as in Chatterji & Tilley (2002) there should, quite properly, be two incentive compatibility constraints.[5]

In summary, if health state is not verifiable the optimal contract in a separating equilibrium consists of:

- two incentive compatibility constraints
- two wages – a wage, w, and sick pay, s, with $w > s > 0$ – that sorts agents according to their type.

[4] Both constraints bind with equality since the principal's objective function is decreasing in w_H, w_S, and w_I.

[5] There are also two pooling equilibria: one in which all agents attend and one in which all agents stay at home. It can be shown that the principal prefers the pooled (all attend) equilibrium to the separating equilibrium as the probability of being ill, ρ, increases.

4 Conclusions

This paper has sought to confront the stylised facts concerning absenteeism with the predictions of two recent theoretical papers in the area. A number of stylised facts of absenteeism in the UK remain unexplained by Skåtun (2003). Indeed a number of its empirical implications are simply not observed.

In particular, there are two specific predictions made by Skåtun (2003) that are of importance in the structure of optimal contracts designed to economise on the costs of absence. First, $w_S > w_H$ implies that absent agents are paid more than healthy agents who present for work. Second, while there is some evidence that firms condition sick pay on the basis of previous, verifiable, absence rates (Brown (1994)) $w_I \neq w_H$ implies that agents who choose to, or are forced to, attend are paid are according to their health state. Neither of these implications is observed empirically. Moreover, any principal structuring their remuneration in this fashion may also encounter adverse selection problems. Indeed, we have shown that only in very limited circumstances can the firm actually structure a wage contract with three components. For the most part, firms will have to seek to manage unwarranted absence and unwarranted presence at work by use of only a two component payment structure. This two component structure – a wage and a sickness pay – corresponds closely to what is observed. The task of the future is to incorporate into the Chatterji & Tilley (2002) two component pay structure, the notion that the threat of epidemics constitutes a further hazard which the firm will seek to manage and control and which may lead to the firm switching from the separating to the pooled equilibrium.

References

Barmby, T. A., Orme, C. D. & Treble, J. G. (1991), 'Worker absenteeism: An analysis using microdata', *Economic Journal* **101**, 214–229.

Brown, S. (1994), 'Dynamic implications of absence behaviour', *Applied Economics* **26**, 1163–1175.

CBI (2002), *Counting the Costs: 2002 CBI Absence and Labour Turnover Survey.*

Chatterji, M. & Tilley, C. (2002), 'Sickness, absenteeism, presenteeism, and sick pay', *Oxford Economic Papers* **54**(4), 669–687.

LMT (2003), 'Sickness absence', *Labour Market Trends* **111**, 65.

Skåtun, J. D. (2003), 'Take some days off why don't you? endogenous sick leave and pay', *Journal of Health Economics* **22**, 379–402.