Employee Absence Costs, Economic Losses, and Firm Competitiveness

Carl P. Kaiser
Washington and Lee University

Abstract: Studies cited in the professional human resource (HR) literature indicate that employee absence costs are very large. However, these studies are not based upon an appropriate theoretical framework for identifying the real losses that may be imposed upon firms by absent workers. The model in this paper incorporates the benefits and costs of absence for both workers and the firm and shows their effects on the equilibrium levels of employment, the number of employees working, and hence, the average number of absences that maximize profits and utility over an annual planning period. The analysis provides a theoretical foundation for distinguishing between absence costs that impose real losses upon the firm and those that do not. The paper concludes with recommendations for firms with regard to correctly assessing absence costs and for efficiently managing absences.
Key words: Absenteeism, unplanned absence, absence costs, absence control

Introduction

Why have several generations of researchers working from a number of disciplines devoted so much time and effort to studying absenteeism? The most obvious answer is that absenteeism is very costly to firms; it would not have been of such great concern if its impact on the firm's performance was not thought to be large. Indeed, the results of studies by professional human resource (HR) service organizations, frequently cited over the last decade, seem to confirm that absenteeism imposes very large costs upon firms. Financial information and employment and attendance records provided by firms and the responses given by corporate officers on absenteeism surveys indicate that the costs of absenteeism are manifested in the following forms: (1) direct costs given by the total of benefits or wages paid to absent workers, (2) indirect costs attributable to lower productivity and/or the additional costs of replacement workers required to cover for absent workers, and (3) administrative costs of keeping records and tracking absences, distributing sick pay, determining and communicating absence policies to employees, and implementing mechanisms to control absences. To evaluate what these costs imply for firm performance we also need to distinguish between planned and unplanned occurrences. While vacations, holidays, and other scheduled absences impose costs upon the firm, they can be handled efficiently by the employer since they are planned. On the other hand, employers find it more difficult to understand and deal with unplanned absences that may impose unexpected losses upon the firm. Following a review of the relevant literature, this paper will provide an economic analysis that clarifies the real consequences of the costs imposed upon firms by unplanned absenteeism.

Absence Cost Measures from HR Studies

The results of a survey by Mercer Health and Benefits LLC, reported in the "Survey on the Total Financial Impact of Employee Absences" (2010) are frequently quoted. Based upon data provided by 276 employers across all major industry segments and regions in the US, this study provided measures of absence costs broken down by the source of the cost (for example, direct payments) and the type of absence (planned or unplanned). The study indicated that the average worker took 5.4 days a year of incidental unplanned absence or approximately 2.1% of total annual workdays. The study also found that the average total cost of incidental unplanned absences amounted to 5.8% of payroll, which was associated with a 19% loss in daily labor productivity (measured as potential lost revenue due to work not covered by replacement labor). Moreover, this survey led to the conclusion that the indirect costs of absenteeism are much larger than the direct costs. Similarly, the ADP Research Institute survey (2012) of 503 employers reported average annual absence rates of 3.2% and 3.5% for midsized and large firms, respectively. The employers also reported large negative impacts of absenteeism on productivity and profitability.

Another well-known study is the "CCH Unscheduled Absence Survey" (2007), which had been carried out annually up to 2007 by CCH—a leading US provider of human resource and employment information. Results from the last survey in 2007 are based upon responses given by 317 HR executives across all major industries and states. This study found an average absence rate of 2.3% calculated as total paid unscheduled absence hours divided by total paid productive hours. The data also clearly identified behavioral patterns indicating that opportunistic workers are more likely be absent on days adjacent to weekends and holidays.

Similar absence patterns have been found by HR firms in other countries, for example, the United Kingdom. The Confederation of British Industry (CBI)/Pfizer absence survey (2011) of UK firms presented results based upon data collected from 223 firms of varying size across a wide range of industries. The study indicated that overall annual absence in the UK in 2010 was 6.5 days per employee at a median cost of £760 per absent employee. The CBI/Pfizer survey (2010) based on data from the prior year (2009) also sought to collect information on productivity losses. But, given the measurement difficulties only one in three respondents were able to estimate such indirect costs. However, a general but soft conclusion of that study was that productivity losses plus other indirect costs stemming from diminished customer service and satisfaction and costs of absence prevention exceed the direct costs of absenteeism.

The results of another British survey of 667 firms by the Chartered Institute for Personnel Development (CIPD) as documented in its annual report on absence management

(2012) indicated an average of 6.8 days of annual unplanned absence per employee (about 3.3% of work time). The median cost of absence per employee for all firms in the survey was reported to be £600 per year. However, this report also indicated that the methods of measuring absence costs were not uniform across the firms responding to the survey where most firms included only the costs of sick pay.

Interpreting the results of these studies is difficult for a number of reasons: First, there is no consistency across these studies in defining the relevant measures. Second, efforts to measure productivity effects have been limited. But, third, and most importantly, these studies lack an analytical foundation for fully understanding the real costs of employee absence; in brief, they assume an unrealistic benchmark of zero absence, and the measurement and interpretation of all absence costs, including productivity costs, have not been sufficiently informed by economic theory. While data from HR studies provide useful information, they do not adequately clarify the economic impacts of absenteeism on the firm's performance.

Academic Analyses of Absenteeism

Before proceeding, a brief summary of the relevant scholarly literature is called for. Absenteeism is a workplace behavior that has concerned social science researchers for many decades. Initially, researchers carried out their work independently, studying absenteeism from the unique perspectives of their own disciplines. Psychologists focused on the behavior of individuals while social psychologists focused on the constraints imposed upon individual behavior within a social context, those working from the perspective of organizational behavior focused on the impacts of organizational structures on the firm's performance, and economists applied the benefit/cost perspective of economics to decisions made by both employers and employees.

To arrive at a comprehensive view of what we knew about employee absence, Kaiser (1998) reviewed and summarized the key findings emerging from several bodies of literature and concluded that they are complementary. As he noted, research on absenteeism exploded in the latter part of the twentieth century with the initial focus primarily upon individual behavior—Why do workers choose to be absent? Most of the management literature on absenteeism was based upon this supply-side approach. But, beginning in the 1970s and 1980s social psychologists and economists began to make important contributions to the literature. Economic perspectives on absenteeism explicitly incorporated the recognition that when workers choose to be absent their decisions are influenced by their environment, which includes the social context in which they work and the economic conditions and constraints that are created by their employers.

In another paper Kaiser (1996) presented a process model explaining work group absence rates similar to the prevailing models in the management literature that emphasized worker characteristics as the prime determinants of absenteeism, but that also incorporated the social context of the work group and the firm's economic and technological constraints. An implication of this model was that absence rates are equilibrium outcomes. The model identified several feedback loops through which the firm may act to reduce future absence rates. This model contributed to the management literature by clarifying that absence rates are endogenously determined by employers and employees.

This interpretation is supported by the equilibrium outcomes predicted by economic models. Allen (1981a) made the initial contribution to this literature by applying Rosen's (1974) hedonic theory of wages to absenteeism. In a competitive world of heterogeneous workers and firms the theory predicts that absence rates should vary systematically across firms and markets; over the long run workers will sort themselves among firms such that those who require the greatest absence rates will be employed by firms that have the greatest tolerance (or incur the lowest costs) for absence. Where the opportunity to be absent is a job characteristic that provides utility, theory predicts that such workers would be willing to accept lower wages than other workers, and firms who face the greatest cost would be willing to pay higher wages for more reliable workers (Allen, 1981a; Weiss, 1985; Coles and Treble, 1993, 1996). Market equilibria should reflect compensating wage differentials for the opportunity to be absent.

Empirical evidence supports the theory. Allen's (1981b) initial empirical analysis confirmed the wage/absence trade-off he predicted. In a third paper Allen (1983) estimated the trade-off between absence rates and output per hour of labor and concluded that the effect of absenteeism on productivity is quite small. Coles and Treble (1993) showed that firms are willing to pay higher wages to reduce the cost of sick pay. Lanfranchi and Treble (2010) investigated the relationship between technology and absence. Where the costs imposed upon firms depend upon the firm's technology, they showed that firms will pay higher wages to acquire more reliable workers if absenteeism is relatively costly. In a related paper Coles, Lanfranchi, Skalli, and Treble (2007) estimated the "cost" of absenteeism of a 1% increase in absence for two technologies. They found a higher shadow price for the technology that is most compromised by a given absence. Finally, using linked employer-employee data, Dionne and Dostie (2007) estimated a small and weakly significant negative correlation between wages and absences, but their results also indicated that alternative work arrangements such as work-at-home options and work schedules are more important determinants of absences than are variations in wages.

These economic models confirm that equilibrium absence outcomes are the result of market forces and that firms respond in ways that are intended to reduce their costs. The model I will develop in this paper also implies equilibrium absence outcomes driven by market forces. But, where we know that some firms are willing to pay compensating wage differentials to reduce absence, we also know that many firms choose other means to manage absence such as closely monitoring workers, creating incentives that alter the absence decision, and/or introducing programs to provide workers with more flexible work schedules. Accordingly, the following equilibrium model assumes a fixed wage rate over an annual planning period. This approach will allow us to decompose absence costs into those that impose real losses upon the firm and those that do not. Thus far there is no documentation of research that has explicitly provided an analytical framework for clarifying the circumstances under which absence costs may or may not impose losses upon the firm that result in diminished competitiveness. The model presented in this paper seeks to fill that gap.

The Model

Workers are assumed to be permanent employees of the firm. Even though they may not show up for work on a given day, employers expect them to return to work after a (typically) short period of time has elapsed. Therefore, we assume that the firm and the worker are parties to an implicit labor contract extending over some period of time, say, one year. The perceived value of such a contract to the individual worker is their expected wage income, expected pay while absent, expected utility of time away from work, and the expected utility from the intrinsic value of work time. The employer anticipates that the average worker will be absent occasionally and is committed to pay the worker for a number of days of absence per year up to a given maximum.

The firm's production function must be specified to account for occasional absences. We assume that the firm's inputs are a fixed capital stock, K, and labor, where in this model two labor variables are required, N_E, the number of employees retained over the year, and N_W, the number of employees who are working on a given day, where the difference between N_E and N_W is the number of workers absent on a given day, N_A. Therefore, the firm's production function becomes

(1) $$Q = Q(K, N_E, N_W)$$

where Q is the firm's daily output rate. However, if some employed workers fill in for others who are absent, then the predicted output will not necessarily be realized; the less substitutable a replacement worker the lower will be their productivity in a given position. Accordingly, the firm's realized daily output, q, is given by $q = (1\text{-}s)Q(K,N_E,N_W)$ where $s = s(N_A)$ is defined as the fraction of potential output that is lost as substitutes fill in for absent workers such that $0 \le s < 1$. If labor is perfectly substitutable, then $s = 0$, but if it is not, then $s > 0$ and is likely to increase with the number of absences such that $s' > 0$. Since the employer is fully aware of the skills required by its technology and should also be aware of the skills possessed by its employees, we assume that s is known with certainty.

Turning to the firm's costs of production, letting r be the price of capital, the firm's daily capital cost then becomes rK. Letting W be the daily wage rate, the firm's wage bill will then be WN_W. Now, suppose B is the benefit amount that a worker receives from the firm for one day of absence. Then the firm would also incur a daily cost of BN_A for those workers who are absent. For purposes of this analysis we will assume that the amount a worker receives for one day of absence is equal to their wage rate. Then the sum of payments made to both those who are working and those who are absent simply becomes WN_E.[1] Finally, additional administrative costs arise that are uniquely associated with the occurrence of absenteeism. When an absence occurs it disrupts the flow of work causing the firm to reorganize work in ways that may entail shuffling job assignments and rescheduling work. Absenteeism also creates administrative costs with regard to the firm's human resource function, which at least entails maintenance of work records and may entail devoting resources to absence control as well. These costs are represented by $C(N_A)$, where $C'(N_A) > 0$.

To investigate the firm's demand for labor when absences are expected over a one-year time horizon we assume that the firm operates in competitive markets. Since competition dictates that a firm's total revenues must be completely exhausted by its total cost, the production function and costs as specified above imply the following constraint:

(2) $$rK + WN_E + C(N_A) = \left[1 - s(N_A)\right] PQ(K, N_E, N_W)$$

Competition also indicates that the utility received from the combination of wage income, sick pay, time absent from work, and the intrinsic value of work time received by the average employee of the firm must be at least as great as the utility they could receive elsewhere; in order to retain its employees the firm must meet their opportunity costs. A competitive labor market will generate the combination of employment, absence, wage income, and sick pay that will maximize the utility of workers subject to the economic constraint shown as equation (2). Therefore, the solution to the model can be found by maximizing the utility of the average worker.

In order to solve the model we specify the average employee's daily utility function as

(3) $$U = U(y,v,i)$$

and define each of its arguments as follows. The daily net income received by the average employee, y, is given by

(4) $$y = (1-t_y)W$$

where t_y is the tax rate on personal income. The second argument of the utility function, $v(N_A/N_E)$, is the non-pecuniary value a worker receives from absence, where $v' > 0$ such that utility increases with the probability of being absent on a given day, *ceteris paribus.* Factors that increase v' include the onset of illness or injury, the emergence of personal needs, and increased preferences for leisure. The last argument, $i(N_W/N_E)$, is the non-pecuniary intrinsic value of time spent at work. If the job is rewarding, $i > 0$ and working another day will increase utility so that $i' > 0$. Therefore, utility increases with the probability of working on a given day, *ceteris paribus*. Anything that increases the value of more work time, for example, greater autonomy or responsibility, will increase i'. On the other hand, if the bundle of non-pecuniary characteristics of a job, on balance, is dissatisfying, $i < 0$ and working another day will decrease utility so that $i' < 0$. In this alternative case, utility decreases with the probability of working another day, *ceteris paribus.* A change in any factor that makes work less rewarding, for example, deteriorating working conditions, will increase the absolute value of i' and decrease utility. Moreover, given that working conditions are known to be very bad in some settings, the range of variation of i is likely to be the greatest in the negative direction.

When markets are competitive the utility maximizing labor contract for employees will be consistent with the profit maximizing outcome for the employer. Therefore, maximizing the utility function with respect to N_E and N_W will result in conditions that identify the profit-maximizing number of employees the firm should retain over its one-year time horizon and the profit-maximizing number of employees the firm expects to work each day.

First Order Conditions

Partial differentiation of the utility function with respect to N_E and N_W implies the following two first-order conditions:

$$(5)\qquad u'(y)\frac{\partial y}{\partial N_E}+i'(N_W/N_E)\frac{\partial y(N_W/N_E)}{\partial N_E}+v'(N_A/N_E)\frac{\partial (N_A/N_E)}{\partial N_E}=0$$

$$(6)\qquad u'(y)\frac{\partial y}{\partial N_W}+i'(N_W/N_E)\frac{\partial y(N_W/N_E)}{\partial N_W}+v'(N_A/N_E)\frac{\partial (N_A/N_E)}{\partial N_W}=0$$

By solving equation (2) for *W* and substituting the result into equation (4), the partial derivatives of *y* with respect to N_E and N_W can be derived. Substituting these partial derivatives and the partial derivatives of N_W/N_E and $(N_E-N_W)/N_E$ with respect to N_E and N_W into the first-order condition then yields the values of marginal product for the last employee retained by the firm and the last of those employees who work as follows:

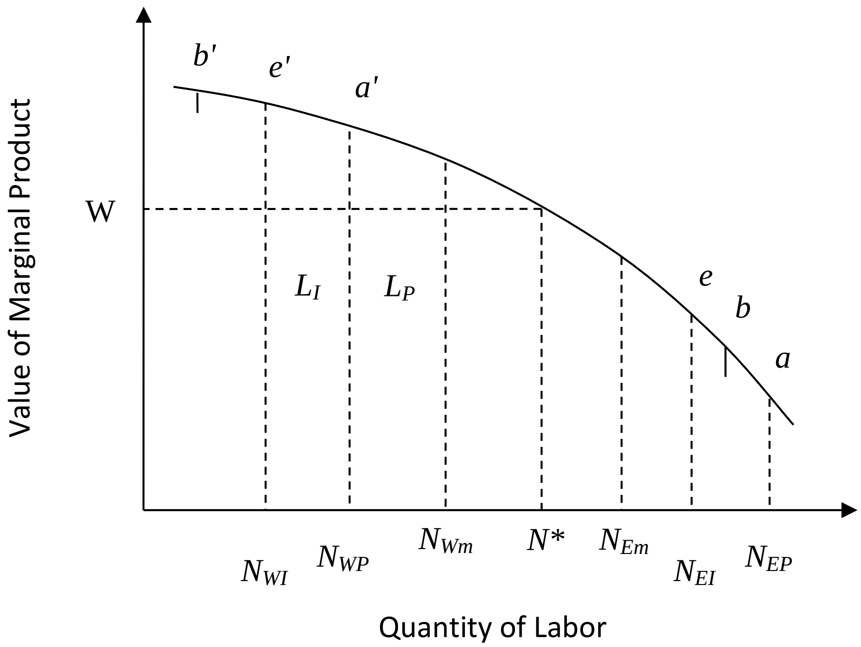
$$(7)\qquad PQ_{N_E}=\frac{1}{\left[1-s(N_A)\right]}\left[W+PQs'(N_A)+C'(N_A)+\frac{i'(N_W/N_E)}{u'(y)(1-t_y)}\frac{N_W}{N_E}-\frac{v'(N_W/N_E)}{u'(y)(1-t_y)}\frac{N_W}{N_E}\right]$$

$$(8)\qquad PQ_{N_W}=\frac{1}{\left[1-s(N_A)\right]}\left[-PQs'(N_A)-C'(N_A)-\frac{i'(N_W/N_E)}{u'(y)(1-t_y)}+\frac{v'(N_W/N_E)}{u'(y)(1-t_y)}\right]$$

To demonstrate the consistency of these results with standard labor theory we first assume perfect substitutability of labor (s = 0, s' = 0). Also, for the average worker, we assume that absence time provides zero utility and that the intrinsic value of time spent at work is zero. In this case there would be no absence implying also that the firm's administrative costs would be zero. With these simplifying assumptions the first order conditions collapse to the simple textbook condition for the profit maximizing quantity of labor: VMP = W. With zero absence the firm employs $N_E = N_W$ workers where the VMP of labor equals the marginal cost of labor, *W*. This outcome is identified as N^* in Figure 1.

But, once we account for the occurrence of absence the firm must choose two optimal quantities of labor—N_E and N_W (or equivalently, N_E and N_A)—as implied by the first order-conditions; the marginal benefits and marginal costs of absence drive a wedge between the optimal number of employees retained over the annual planning period and the optimal number of employees who are expected to attend work on a given day. First, note that the wage rate, *W*, appears in equation (7) but not in equation (8). This is as expected since the wage rate is the fundamental determinant of the marginal cost of labor with regard to the number of workers to employ while N_W is independent of the wage rate, given N_E.

Figure 1. Equilibrium Employment and Absences



Inspection of the terms appearing in the first order conditions reveals their obvious symmetry. First, consider the terms involving *v'* that capture the effects of the non-pecuniary value that workers receive from being absent. These terms measure the monetary value of the marginal benefit a worker receives from being absent in terms of equivalent pre-tax earnings. Since *v'* appears with a negative sign in equation (7) and a positive sign in equation (8), the model implies that an increase in the marginal utility of absence will increase N_E, decrease N_W, and therefore increase the number of workers absent on a given day.

Similarly, the terms involving *i'* capture the effects of the non-pecuniary, intrinsic value associated with the characteristics of the job. These terms measure the monetary value of the marginal benefit a worker receives from working in terms of equivalent pre-tax earnings. The model implies that characteristics of a job may increase or decrease absence. First, if the content of the job provides the worker with an intrinsic reward (in addition to wage income), then $i > 0$ and $i' > 0$. Since the sign on this term is positive in equation (7) and negative in equation (8), an increase in the marginal benefit of work time will decrease N_E, increase N_W, and consequently reduce the number of workers absent on a given day. This outcome implies that workers respond to the "pull" of rewarding job content and working conditions, *ceteris paribus*. However, we can imagine situations in which working conditions (for example, adversarial and destructive labor-management relations) are such that $i < 0$ and $i' < 0$. In this case the signs on the terms are reversed and therefore model implies a greater number of absences on a given day. In this scenario adverse conditions "push" workers into absence, *ceteris paribus*.

Finally, *C'*, the marginal administrative cost of one absence, appears in equation (7) with a positive sign and in equation (8) with a negative sign. The model consequently implies that an increase in the marginal administrative cost of absence, as expected, decreases the number of workers absent on a given day by decreasing N_E and increasing N_W.

To summarize, given the assumption of competitive product and labor markets, this model explains the amount of absence a firm experiences as an equilibrium outcome. Partial differentiation of the first order conditions with respect to *v′*, *i′*, and *C′* yields the following comparative static implications for absence—*ceteris paribus*, the model implies that on a given day:

1. An increase in the marginal utility of absence time will "pull" employees away from work and increase absence.
2. An increase in the marginal utility derived from working will "pull" employees into work and decrease absence.
3. An increase in the marginal disutility of time spent at work will "push" employees away from work and increase absence.
4. An increase in the firm's marginal administrative costs of absence will decrease absence.

Costs of Opportunistic Absence with Perfect Substitutability of Labor

To proceed, we first assume that $s = 0$ and $s' = 0$ such that $(1-s) = 1$ and $PQs' = 0$ in equations (7) and (8). Accordingly, in Figure 1 the equilibrium outcomes of N_{EP} and N_{WP}, in reference to N^*, are shown as points *a* and *a'* on the value of marginal product curve. As a first step in determining economic losses attributable to unplanned absences we must clarify the appropriate benchmark against which to compare the number of absences the firm experiences. While N^* in Figure 1 shows the number of workers the firm would employ if absences did not occur at all, it is not the appropriate benchmark for determining the firm's losses attributable to absenteeism. For real firms there is some minimum amount of incidental unplanned absence due to unavoidable causes including incapacitating illness or injury and other personal matters. This outcome can be represented by N_{Wm} and N_{Em} in Figure 1. The firm should therefore perceive $N_{Em} - N_{Wm}$ as its lowest possible number of absences, and hence, $(N_{Em} - N_{Wm})/N_{Em}$ as the lowest possible absence rate it can expect, where subscript *m* is used to index the "minimum" absence outcome.

If we assume that the firm's absence rate is at the minimum, then what are the costs imposed upon the firm? Workers who are unavoidably absent will receive sick pay from the employer that replaces wages lost due to absence. But since this is the best outcome the firm and all of its competitors can expect, the sick pay it provides for the absent workers cannot be considered a loss for the firm. Given that workers are employed over an annual planning period, benefits paid for unavoidable absence must be considered necessary costs of production. These benefits cannot be avoided and must be paid to workers in order retain their services over the long run. Similarly, any costs incurred by the firm for administration of minimum absence must be considered a cost of production. Therefore, while the minimum amount of absence imposes costs upon the firm, since the firm's competitors bear these costs as well, they do not constitute economic losses or diminish the firm's competitiveness. Moreover, if we initially assume that the probability of occurrence of unavoidable absence is uniform across all firms, then the minimum absence rate will be the same for all firms.

How then do we interpret a greater equilibrium number of absences as illustrated by points *a* and *a'* in Figure 1? The typical approach of the HR studies cited above is to measure the direct cost of absenteeism as the sum the direct payments made by the firm to all absent workers. In terms of the model shown in Figure 1, the direct costs of absenteeism recorded from each firm in such studies would be the total sick pay given to N_{EP} - N_{WP} workers. As these HR studies are carried out and presented, the inference is that all of these direct costs are damaging to the firm. However, to arrive at a measure that will better allow us to identify any economic losses that may be incurred, an estimate of how much a firm would pay in benefits if the absence rate was at its minimum must be subtracted from total amount that is paid. This is the measure that indicates the costs that may be imposed upon the firm by opportunistic unplanned absence.

With regard to the administrative costs of absenteeism, those that are associated with minimum absence are, again, necessary and unavoidable costs for all firms and should not be seen as imposing a competitive disadvantage on the individual firm. Therefore, figures that measure the total costs of absence administration should also be taken as overestimating the relevant costs imposed upon the firm. Here, again, it is useful to decompose administrative costs into those that are necessarily incurred with minimum absence and those additional costs that result from opportunistic absences. Accordingly, conventional estimates of administrative absence costs are also overstated.[2]

Where the HR studies cited above simply offer estimates of diminished productivity, output per unit of labor time, the model developed here provides an explanation of how opportunistic absences impact upon production. Returning to the minimum amount of absence, Figure 1 shows N_{Wm} as the maximum number of workers, on average, who will attend work on a given day. Therefore, the output rate generated by this many workers, Q_{Wm}, becomes the maximum average daily output rate that the firm can achieve. If more than the minimum number or workers are absent, then the firm's output rate must be less than Q_{Wm}. If, in general, we take the number of employees working to be N_{WP} as shown in Figure. 1, then the value of the output lost due to opportunistic absenteeism is shown by the area labelled L_P up to the VMP curve. Importantly, this result implies that *opportunistic unplanned absence imposes an output loss upon the firm, even when the firm's employees are perfect substitutes for each other*. Because absence increases the firm's marginal cost of production, its profit-maximizing output falls.

Costs of Opportunistic Absence with Imperfect Substitutability of Labor

In general, those who fill in for absent workers on a given day are not likely to be perfect substitutes; the productivity of a replacement will be less than that of the worker who regularly fills that position. Moreover, if the productivity of other workers depends in some way upon the performance of the worker filling this position, then the output of other workers in a unit may also be reduced. In the above analysis we determined the costs attributable to opportunistic absenteeism for the case in which $s = 0$ and $s' = 0$. Given these results, we can now show how imperfect substitutability of labor will impose additional costs upon the firm.

Where (1-*s*) enters the model as a coefficient of *Q*, higher values of *s* represent less substitutability of replacement labor and imply less realized output for a given quantity of labor. In terms of equations (7) and (8), the effect of imperfect substitutability ($s > 0$) can be interpreted as increasing the marginal cost of labor since (1-s) appears in the denominator of the marginal cost side of both equations.

To further identify the effects of imperfect substitutability on firm costs we begin with a simple, intermediate case. If we first make the simplifying assumptions that v', i', and C' are constant and also that $s' = 0$, the first order conditions then imply that that the equilibrium values for N_E and N_W would be given by points b and b' as shown on the VMP curve in Figure 1. However, assuming that $s' > 0$, where PQs' enters equations (7) and (8) with opposite signs, points e and e' become the equilibrium values of N_{EI} and N_{WI}, respectively, that capture the primary impact of imperfect substitutability.

While this gives us the primary and dominant impact of imperfect substitutability, we now relax the restrictive assumptions made on v', i', and C' to show its full impact. The primary impact of imperfect substitutability on the number of absences will create feedback through the last three terms of equations (7) and (8). But in making the standard assumptions of diminishing marginal utility and rising marginal cost such that $v'' < 0$, $i'' < 0$, and $C'' > 0$, these secondary effects are seen to be offsetting and are most likely to be small. Therefore, we take points e and e' as representative of the equilibrium values of N_E and N_W (N_{EI} and N_{WI}) that fully account for the impact of imperfect substitutability. The model then implies that imperfect substitutability reduces both the number employees and the number of those who work as compared to the equilibrium obtained for perfect substitutability of replacement labor, where the result for the number of absences is ambiguous and depends upon the size of s'. However, where substitutability is very limited s' will be relatively large and consequently s will increase quickly, even for small increases in the number of absences. With relatively large values of s' working through equations (7) and (8), the model therefore implies that very limited substitutability of labor will lead to less absence as well lower employment and fewer employees working. This makes intuitive sense; if other employees who fill in for absent employees are not productive in those positions, then there is no need for the firm to retain more employees to cover for absent workers.

To summarize the effects of imperfect substitutability, partial differentiation of the first order conditions with respect to s and s' yields two additional comparative static results for absence—*ceteris paribus*, the model implies that on a given day:

5. A decrease in substitutability of replacement labor (a higher value of s) reduces both total employment and the number of employees who work.
6. An increase in the rate at which substitutability diminishes with more absences (a higher value of s') reduces the number of absences and the absence rate.

Now we are able to identify the additional output losses created by absenteeism when substitutability of labor is imperfect. Total output will be lower because fewer employees will be working; $N_{WI} < N_{WP}$. Therefore, the output loss incurred by the firm due to imperfect substitutability is shown as the area labelled L_I up to the VMP curve in Figure 1. Finally, the total output loss incurred by the firm due to opportunistic unplanned absences is therefore given by the sum of areas L_P and L_I up to the VMP curve.[3] These results unambiguously imply that output losses due to opportunistic unplanned absenteeism are largest when the ability to substitute labor is least. Imperfect substitutability substantively increases the firm's marginal cost of production thereby causing a leftward shift of its product supply curve and a further drop in daily output.

Given worker preferences, the firm's technology, product demand, and the benefits and costs of absenteeism perceived by both the employer and employees, the equilibrium outcomes modelled above will achieve maximum utility for workers and maximum profit for firms. With imperfect substitutability and the occurrence of opportunistic absences the profit-maximizing firm will employ N_{EI} workers such that the profit-maximizing output will be produced by N_{WI} workers—the average number of workers the employer expects to attend on each workday over the planning period. Despite the difficulties of measuring productivity losses as gleaned from management observations, the general conclusion implied by the HR studies discussed above is that productivity losses are very large. This is observation is consistent with the theoretical prediction of the model presented in this paper, particularly with imperfect substitutability of replacement labor.

Absenteeism and Firm Performance

Absenteeism diminishes the firm's profit, first, by increasing the "costs" it incurs in dealing with absenteeism, and second, by reducing its "total revenue" as a consequence of the impact of absenteeism on the firm's output rate. Where a given firm is in competition with other firms its greatest concern should be for losses imposed by costs of absenteeism that are not incurred by other firms. To correctly assess its competitive position, a firm needs to compare its absence outcome to the appropriate benchmark. But what is the appropriate benchmark?

Given that different products are produced using unique production functions that entail different labor requirements, in competitive circumstances we observe different average absence rates across industries.[4] Where firms and workers within a given industry are not homogeneous we expect firm absence rates to be distributed around an average rate observed for the industry. Since the average absence rate observed in a competitive industry is determined by market forces, it therefore becomes the absence rate critical to understanding economic losses that may be incurred by a given firm within the industry. We therefore identify the industry absence rate as the benchmark absence rate for evaluating a given firm's competitiveness.

The equilibrium quantities of labor for a representative individual firm when both unavoidable and opportunistic absences occur are shown as N_{EI} and N_{WI} in Figure 1. This outcome for a given firm depends upon the exogenous variables in the first order conditions. If we assume that v' (marginal value of absence time) and i' (marginal value of the intrinsic reward for work time) may vary across firms, the model implies that the amount of opportunistic absence will vary across firms. Worker incentives to be absent for opportunistic reasons are stronger with higher values of v' and/or lower values of i'. As implied by the comparative static results 1, 2, and 3 listed above, with greater values of v' and/or lower values of i', (1) N_{EI} will be greater, (2) N_{WI} will be lower, hence, (3) more workers will be absent, and (4) the absence rate will be greater.

Now, consider the following two alternative outcomes: If, in the short run, a single firm in an industry experiences an absence rate over its planning period that is greater than the benchmark for the industry, the firm will be incurring economic losses that reduce its ability to compete with other firms. *Ceteris paribus*, the equilibrium absence rate for such a firm minimizes its economic loss, given the technology, organizational structure, human resource policies and procedures chosen by the firm, and the preferences and needs of its workforce. This outcome is most likely to occur for firms that do not adequately understand the determinants of absenteeism and whose efforts to manage absenteeism are ineffective and/or insufficient or non-existent.

On the other hand, if a firm's equilibrium absence rate is less than the industry benchmark absence rate, *ceteris paribus*, the cost incurred by the firm due to opportunistic absences is less than that incurred by most firms in the industry, and therefore it will enjoy a competitive advantage over most other firms; in the short run this firm will earn an economic profit relative to the a normal profit earned, on average, by other firms in the industry.

To summarize, the above analysis addresses two types of unplanned absences; opportunistic and unavoidable. First, with regard to opportunistic absences, competition creates incentives for firms incurring economic losses to introduce changes that will eliminate excessive absences. If such a firm does not succeed, then, *ceteris paribus*, it will be compelled to shut down in the long run. Second, competition also creates incentives for firms to address the causes of unavoidable absence in order to devise ways of reducing it. Here we relax the prior assumption that minimum absence rate is the same across all firms. All firms can gain by introducing changes that lower the rate at which employees are unavoidably unable to attend work, for example, wellness programs, more flexible work arrangements, more rewarding job content, or more substantive and clear discipline for excessive absences. Whether a firm's absence rate is above or below the benchmark for the industry, the firm can improve its competitiveness by taking such actions. In terms of the model, reduced rates of both opportunistic and unavoidable types of absence would be accomplished by taking actions that reduce the value of v' and/or increase the value of i' as in equations (7) and (8). For example, provision of daycare and flexible work times are likely to reduce the demand, and hence, the utility of absence, while the intrinsic utility of a day at work is likely to increase with improvements of working conditions .[5]

Implications for Management

There are a number of implications of this analysis for employers. First, the employer should understand that the placement of workers within firms is the result of market processes that lead to equilibrium outcomes. Workers who bring their skills, preferences, and needs to the labor market interact with employers who demand workers who have the skills dictated by their production functions and who have strong commitments to the organization and its goals. This analysis implies that every firm will observe a unique absence rate that depends upon the mix of characteristics embodied in its workforce and the mix of job situations created by the firm. To illustrate, if the firm's workforce, on average, receives higher utility from time spent away from work than workers employed by other firms, then, *ceteris paribus*, market forces will assure that this firm will have higher absence rates than other firms. Similarly, if a given firm provides extraordinarily good working conditions, *ceteris paribus*, it will have less absenteeism than other firms. Or, absence rates will be lowest where production is carried out with the highest skilled and least substitutable workers, *ceteris paribus*.

Second, employers should identify the causes of the absences taken by their employees. With this knowledge in hand, a critically important practical strategy for a firm to follow is to determine a "target" absence rate. Obviously, this rate should be fundamentally based upon unavoidable absences that result from unambiguous inability to attend work. But, in practical terms, a perfect distinction between ability and inability to attend is not possible. Therefore, employers should seek to determine and communicate to workers what they consider to be legitimate causes of inability to attend. For example, the firm may set a target absence rate that incorporates personal reasons that create a need to be absent such as caring for a sick child. Where it's clear that there is some discretion over what a firm should include in its target absence rate, the firm must carefully record and track unplanned absences by employee and reported cause as they occur over a long period. The period should be long enough to generate average absence rates by cause that will inform a target absence rate for future planning periods.[6]

Third, once a target absence rate has been determined based upon data collected over a sustained period of time, using the difference between the target and the actual absence rates, the firm would be able to arrive at an estimate of the real cost of opportunistic absence it is incurring. Where the rational firm would compare the benefits and costs of each potential use of its scarce resources, having a reasonably good estimate of the real cost of absenteeism is obviously important to the firm with regard to determining the optimal amount to invest in its efforts to manage employee absences.

Fourth, if the firm has information on the average absence rate for all the firms in its industry, by comparing its own absence rate to the industry average, it can determine whether its absence rate confers upon it a competitive advantage or imposes upon it a competitive disadvantage relative to the average firm in the industry.

Fifth, if the firm wishes to bring its absence rate down to it target rate, it should implement programs and policies intended to reduce the propensity of its employees to choose absence on a given day. Such policies may be positive or negative; they may reward workers for excellent attendance or they may discipline employees who take unplanned absences at high rates. Similarly, if firms implement supportive policies such as flexitime that allow workers to manage time more efficiently, the need to take absence (value thereof) will be reduced. While these efforts can lead to reductions in absence over a short run time horizon, managers should understand that these improvements can only be marginal. In other words, with a given technology and a given work force, there are limits to how much equilibrium absence rates can be influenced. Over the long run, the firm's choice of technology and its hiring standards must be considered in any efforts to achieve lasting and substantial reductions in its absence rates.[7]

Conclusions

While conventional studies carried out by human resource service organizations provided useful information for employers in terms of measuring gross costs of absenteeism, the model developed in this paper demonstrates that these estimates are not conceptually well grounded for the purpose of assessing the impacts of absenteeism on firm performance. The results these HR studies present are based upon accounting and financial data kept by firms for purposes other than deriving accurate estimates of the economic costs of absenteeism. They implicitly assume that the observed employment of a given firm is the correct reference point for studying absenteeism—in effect, a benchmark of zero absences. This approach neglects that the behaviors of workers and firms are driven by market forces that lead to equilibrium outcomes, including the absence rates we observe. Therefore, these gross figures fail to differentiate between what are fundamentally costs of production and real economic losses and gains that correspond to a distribution of absence rates across competitive firms.

The model presented here demonstrates that employment levels and absence rates are endogenously determined implying that observed employment levels cannot be taken as a benchmark level of employment. To arrive at estimates of the real losses that may be imposed by opportunistic unplanned absences, the appropriate benchmark is the outcome the firm would attain if its absence rate was equal to its industry average. The model therefore indicates that the direct and administrative costs of absenteeism that are critical to efficient decision making by firms are lower than those suggested by the HR studies.

Moreover, as noted in the introduction, efforts to measure productivity losses have been very limited, and when it has been attempted, it has been done in various ways, none of which have been very satisfying. The model developed in this paper implies that output losses can be large, even when replacement labor is perfectly substitutable for absent labor. Once a firm has a reasonable estimate of its minimum absence rate and then sets a target absence rate, and consequently identifies the maximum average number of workers it can expect to attend on a given day, then it should be able to arrive at an estimate of its potential output. An estimate of the output losses due to absenteeism can then be found by comparing actual output to potential output. While the result, by its nature, is an estimate rather than a precise measure, it rests on a sound theoretical foundation, which cannot be said of current perceptions of losses due to absence. If employers wish to fully understand the consequences of unplanned absence and design the most efficient programs and policies for managing absence, their efforts need to be grounded in a conceptually sound framework of analysis.

Notes

1. Since this model assumes that the firm privately administers both the financing and payment of absence benefits to its own employees, this analysis is most directly relevant to US firms rather than firms in other OECD countries where absence benefits are administered through national sickness absence programs.
2. This model focuses on the average daily absence rate expected over an annual time horizon and so does not specifically deal with the costs of hiring temporary replacement workers and overtime payments, which firms frequently cite as contributing to the direct costs of

absenteeism. However, on any given day greater or fewer than the expected number of absences may occur. Given that the average rate is observed over the firm's time horizon, in the context of this model the costs of hiring and overtime that may occur during spells of high absence will be at least partially offset by the relatively low sick pay costs incurred during spells of abnormally low absence.

3. One qualification needs to be made regarding the output loss due to imperfect substitutability shown in Figure 1. Following the logic of the equilibrium outcome, the minimum number of employees who work on a given day must fall slightly as well. Therefore, the total output loss when labor is less than perfectly substitutable is slightly less than the area $L_P + L_I$, but clearly larger than the loss L_P as shown in Figure 1.
4. See U.S. Bureau of Labor Statistics for 2012 absence rates by industry (2013). Chadwick-Jones, et al. (1982) in their work on the social psychology of absenteeism, were the first to suggest that these differences are systematic.
5. Taking such actions will increase the administrative costs of absenteeism, but they are likely to be more than offset by the lower direct and indirect costs as a result of lower absence rates.
6. Given the much observed pattern of higher absence rates on days adjacent to weekends and holidays, absence rates for these days should obviously not be used to identify the target rate since they clearly occur for opportunistic reasons.
7. Kaiser's (1996) model illustrates the feedback processes through which management actions can reduce absence rates in both short run and the long run.

References

ADP Research Institute. (2012) *Total Absence Management: Two Decades after the Passage of FLMA*. ADP, Inc., Roseland, NJ.

Allen, S.G. (1981a) Compensation, safety, and absenteeism: Evidence from the Paper Industry. *Industrial and Labor Relations Review*, 34: 208-218.

Allen, S.G. (1981b) An Empirical Model of Work Attendance. *Review of Economics and Statistics*, 63: 77-87.

Allen, S.G. (1983) How much does Absenteeism Cost? *The Journal of Human Resources*, 18: 379-393.

CCH, Inc. (2007) *CCH Unscheduled Absence Survey*. Riverwoods, IL.

Chadwick-Johns, J.K., Nicholson, N., & Brown, C. (1982) *Social Psychology of Absenteeism*. New York. Praeger.

Chartered Institute of Personnel and Development. (2012) *CIPD Annual Survey Report on Absence Management*. London.

Coles, M., & Treble, J. (1993) The Price of Worker Reliability. *Economics Letters*, 41: 149-155.

Coles, M., & Treble, J. (1996) Calculating the Price of Worker Reliability. *Labour Economics*, 3: 169-188.

Coles, M., Lanfranchi, J., Skalli, A., & Treble, J. (2007) Pay, Technology and the Cost of Worker Absence. *Economic Inquiry*, 45: 268-285.

Confederation of British Industry. (2010) *CBI/Pfizer Absence and Workplace Health Survey*. London.

Confederation of British Industry. (2011) *CBI/Pfizer Absence and Workplace Health Survey*. London.

Dionne, G., & Dostie, B. (2007) New Evidence on the Determinants of Absenteeism using Linked Employer-Employee Data. *Industrial and Labor Relations Review*, 61: 108-120

Kaiser, C.P. (1996) Individual, Social, and Economic Determinants of Employee Absence: An Integrative Analysis. *Journal of Business and Economic Studies*, 3: 1-20.

Kaiser, C.P. (1998) What do We know about Employee Absence Behavior? An Interdisciplinary Interpretation. *Journal of Socio-Economics*, 27: 79-96.

Lanfranchi, J., & Treble, J. (2010) Just-in-Time Production, Work Organisation and Absence Control. *The Manchester School*, 78: 460-483.

Mercer (US), Inc. (2010) *The Total Financial Impact of Employee Absences*. New York.
Rosen, S.S. (1974) Hedonic Prices and Implicit Markets. *Journal of Political Economy*, 82: 34-55.
U.S. Bureau of Labor Statistics. (2013) *Labor Force Statistics from the Current Population Survey*. Table 47, Absences from Work of Employed Full-Time Wage and Salary Workers by Occupation and Industry.
Weiss, A.M. (1985) Absenteeism and Wages. *Economics Letters*, 19: 277-279.

Contact Information:

Carl P. Kaiser
Department of Economics
Washington and Lee University
Lexington, VA 24450
kaiserc@wlu.edu
(540)458-8617