This is a self-archived version of an original article. This version may differ from the original in pagination and typographic details.

Author(s): Böckerman, Petri

Title: Presenteeism in Economic Research

Year: 2018

Version: Accepted version (Final draft)

Copyright: © Cambridge University Press, 2018.

Rights: In Copyright

Rights url: http://rightsstatements.org/page/InC/1.0/?language=en; http://rightsstatements.org/page/InC/1.0/?language=en; http://rightsstatements.org/page/InC/1.0/?language=en

Please cite the original version:
Presenteeism in economic research*#

* Turku School of Economics, Labour Institute for Economic Research and IZA. Address: Pitkänsillanranta 3A, FI-00530 Helsinki, Finland. Phone: +358-9-2535 7330. Fax: +358-9-2535 7332. E-mail: petri.bockerman@labour.fi

# Work related to this article has been financially supported by the Strategic Research Council funding for the project Work, Inequality and Public Policy (293120). I am grateful to Markku Jokisaari and Terhi Maczulskij for useful comments regarding the earlier version of this article. The usual disclaimer applies.

Introduction

Change is a constant ingredient of modern work life. Thus, job demands and the complexity of job tasks that workers are required to accomplish and perform in everyday work are increasing in organizations and firms. Increasing job demands and incessant changes may be stressful, at least for certain employees, and cause significant and long-lasting health effects as well. Most individuals spend a substantial portion of their total time at work. Thus, work-related behaviours have significant spillover effects on other aspects of life. For these reasons, work-related sickness is also a particularly important aspect of employee well-being from the broader economic perspective.

Theoretical considerations in economic research have put special emphasis on the determination of the optimal amount of work-related sickness behaviour. An important feature of formal economic models is the asymmetric information that prevails between
employees and employers. This underlying structure implies that employers cannot directly
observe employees’ health endowment and/or health status, which are the key determinants
of work-related sickness behaviour. This implication makes it challenging for firms to
design optimal policy responses to work-related sickness.

There are two equally important and widespread aspects of work-related sickness
behaviour. These aspects are absenteeism (“absent from work because of sickness”) and
presenteeism (“present at work despite sickness”). There is a large theoretical and empirical
body of literature in economics regarding the determinants and consequences of
absenteeism (Treble and Barmy, 2011). This research has evaluated the optimal
compensation rate of sick leave and substantially broadened the understanding of the
determinants of absenteeism, for example, by examining the effect of business cycle
fluctuations on sickness absence. However, only very recently has theoretical and empirical
research in economics more deeply examined the underlying causes and implications of
presenteeism. Thus, studies from other disciplines such as organizational and health
psychology, and management have examined presenteeism from different angles, well
before economics.

Presenteeism has significant implications for multiple aspects of work life. Presenteeism
costs individual workers, firms and society as a whole. For example, there is a substantial
amount of empirical evidence showing that working while sick leads to worse health
outcomes in the long run (Skagen and Collins, 2016). Thus, working while sick may
contribute to the incidence of disability pension over the long observation period. These
costs are carried by the society as a whole. Ultimately, presenteeism may considerably
contribute to workers’ ill health and firms’ costs. Therefore, it is crucial to have a greater
knowledge of the economic determinants of both sickness absenteeism and presenteeism, which constitute the two important work-related sickness categories.

The objective of this chapter is to provide a condensed overview of both theoretical and empirical work in economic research on presenteeism. For brevity, the discussion is focused on the most important aspects in the emerging literature in economics on presenteeism. Because the chapter focuses solely on research in economics, the number of approaches and studies that are described are very limited.

The chapter is structured as follows. First, we describe the formal theoretical considerations that additionally elucidate the optimal amount of presenteeism. The models consider that employers and employees strive to maximize profits and the perceived utility. We then highlight the key empirical patterns from the sparse empirical literature in economics on presenteeism. The last section of the chapter concludes by stating certain important gaps and challenges in the current theoretical and empirical knowledge of presenteeism from the economics perspective. The concluding section also discusses the specific topics that the literature should put more emphasis on to provide more practical implications that would benefit firms in the design of policies to minimize the costs associated with presenteeism and absenteeism. These lessons could be implemented in organizations and firms to achieve the optimum amount of absenteeism and presenteeism.

Theoretical considerations

Employers and employees are the relevant agents who determine the rate of absenteeism and presenteeism in organizations and firms. Employers and employees respond to the
financial incentives and constraints that are established by sickness insurance systems. Economic theory assumes that both employers and employees make rational choices subject to constraints.

The optimal amount of absenteeism in organizations and firms is clearly not zero for a number of reasons. An important reason for this is that sick workers cause output losses in firms, as discussed in detail below. This feature is arguably particularly important in the context of increasingly popular team work where the total output of the team is often dependent on the performance of the weakest link in the team. Conversely, self-management teams may consist of workers who are multi-skilled, which makes them relatively easy to substitute for each other.

Sickness insurance has significant effects on the financial incentives to work while being sick or to be absent. Sickness insurance is an important element of social insurance in nearly all industrialised countries. Workers are entitled to compensation when they are absent from work due to a certified sickness. Compensation is nearly always proportional to the wage level. Details of the sickness insurance systems such as the maximum amount of compensation vary from country to country. Importantly, there is a moral hazard related to compensation that is provided by sickness insurance because higher compensation levels tend to lead to longer sickness absence spells, all else being equal (Ziebarth and Karlsson, 2014).

An important cost component related to work-related sickness is caused for firms by the fact that absenteeism often leads to the use of replacement workers. Other workers may also have to work paid overtime to compensate for the lost labour input of those who are absent.
Importantly, an increase in paid overtime causes substantial costs to firms because firms must pay an overtime premium. Collective labour agreements typically stipulate the amount of overtime premium; in addition, there is a maximum amount of overtime that a worker can perform during a certain time period. Binding collective labour agreements are particularly relevant in the context of blue-collar occupations. For white-collar workers, in most settings, it is very difficult to measure the output of a single worker in an organization. This statement implies that it is challenging to evaluate the amount of work needed to cover absent white-collar workers in terms of the required overtime. Therefore, his/her lost input due to absence may be covered by co-workers’ unpaid overtime if the lost labour input is needed to achieve the binding production requirements.

Economic incentives also determine the optimal rate of presenteeism. From the employer’s perspective, the standard neoclassical model of firm is based on the notion that a firm combines a set of inputs to produce output. The most common production inputs in the standard model are labour and capital. The fact that certain employees work while being sick is a potentially important determinant of output because effective labour input has two components: hours of work and productivity (output per hours of work). A worker’s current health status and working capacity are the key determinants of output per hours of work. Maximum output per hours of work is possible to achieve only if workers are completely healthy at work and thus contribute fully to the production of output. Importantly, working while being sick may also increase the rate of serious work-related accidents at work, which has significant negative consequences on other workers at the workplace. Therefore, working while sick implies that the output per hours of work is lower than the maximum because of the reduced effort and/or mistakes that working while being sick causes in
organizations and firms, all else being equal. This finding leads to losses in firm’s output and causes substantial costs for firms.

From the employee’s perspective, the standard neoclassical textbook model in economics assumes that individuals only consider their personal, individual-specific costs and benefits while making work-related decisions such as whether to work while being sick or not. Workers work while being sick if their personal utility level is higher than the utility level while being absent from work, all else being equal. This finding is arguably a very limited view.

Importantly, the existence of externalities is a decisive aspect of presenteeism. Externality is the cost or benefit that has an influence on an agent who did not choose to incur that cost or benefit. Importantly, returning prematurely to work from sick leave causes potentially significant negative externalities to co-workers in the workplace. The negative externalities are particularly important in the context of contagious diseases (Pichler and Ziebarth, 2016). This pattern is caused by the observation that working while sick will increase the prevalence of absenteeism in the organization as a whole, which increases a firm’s costs and leads to the reduction of output. Therefore, the evaluation of the existence of negative externalities on co-workers is essential to determine the optimal length of sickness absence.

In addition to negative externalities that are inflicted on co-workers in the workplace, there are also potentially important negative externalities that impact firm’s customers and other relevant agents. For example, in hospitals and other healthcare settings, those who work while being sick may cause substantial harm to patients by spreading contagious diseases.
and may prolong the patients’ recovery from the illness for which they were originally admitted to hospital. Thus, if they are relevant negative externalities inflicted on the firm’s customers, these effects should be fully considered in the determination of the optimal rate of absenteeism. In other contexts, negative externalities arguably have only relatively small, second-order effects that can be ignored in the economic analysis. For example, certain white-collar occupations such as the work of a reporter are very independent, and they do not require day-to-day interaction with co-workers, which would cause substantial negative externalities.

Considering the negative externalities leads to longer optimal sickness absence spells than the standard model, which is solely based on the individual’s utility. However, it is also important to note that the role of negative externalities is not identical for all diseases. The negative externalities are particularly important for contagious diseases such as seasonal infectious diseases, while the negative externalities caused by chronic conditions such as back pain are most likely minor in practical workplace contexts. This finding complicates the considerations regarding the optimal sickness leave scheme because the compensation rate of sickness insurance cannot be conditioned on the exact nature of an illness. Ultimately, the negative externalities to co-workers should be fully considered in the firm’s optimal policy regarding work-related sickness.

There are only three formal theoretical models of presenteeism that have been presented in the economic literature. The early theoretical models include those by Chatterji and Tilley (2002) and Brown and Sessions (2004). These models are based on the notion that employers can fully regulate the optimal amount of presenteeism by setting the wage level and/or specific details of sick pay scheme for the affected workers.
Chatterji and Tilley (2002) argue that there is a particularly strong incentive for employers to offer higher compensation than the statutory sick pay set by sickness insurance legislation. These additional payments are able to prevent ill workers with reduced job performance from being at work because workers respond strongly to financial incentives. Thus, increased sick pay implies that employees remain off of work longer. Hence, by using additional payments, firms can avoid the output losses that are caused by those who work while being sick. Consequently, in the (long-run) equilibrium of the model, there is no presenteeism because firms can fully and freely optimize the amount of output that they produce in the market. Thus, the most important implication of the model is fundamentally in conflict with the empirical evidence on presenteeism.

Based on the augmented efficiency wage model, Brown and Sessions (2004) argue that, by taking advantage of the eligibility criterion for firm-financed sick pay, employers can cause workers to either engage in absenteeism or presenteeism. A firm’s ultimate goal is to reduce the costs associated with presenteeism. Importantly, the model presented by Brown and Sessions (2004) does not allow for the simultaneous occurrence of absenteeism and presenteeism. Additionally, the model assumes that firms have full command over the specific compensation rules of sickness insurance. This statement is clearly not the appropriate characterization of the sickness insurance system in the countries that have statutory sickness insurance. Firms only rarely provide firm-specific insurance policies for workers in Europe. Most European countries have compulsory sickness insurance set by legislation that covers the population. In contrast, in the United States, most firms are directly involved in the design of sickness insurance schemes because insurance policies are provided by employers as part of the compensation package for workers. Taken together, the model by Brown and Sessions (2004) is not able to provide a universal
explanation for the existence of presenteeism that has been reported for all industrialized countries because it focuses on firm-specific aspects of sickness insurance schemes. Thus, the key empirical implications of the theoretical models by Chatterji and Tilley (2002) and Brown and Sessions (2004) are not consistent with the empirical observations that have been convincingly documented in the literature.

Hirsch et al. (2017) present the first complete economic model of work attendance that incorporates both work-related sickness categories (i.e., absenteeism and presenteeism). In the model, presenteeism is interpreted as the outcome of the optimizing behaviour of both employers and employees. An important aspect of the model is that there is asymmetric information between employees and employers. Employees know their underlying productivity and health status, but this information is not directly observable to employers. Thus, health status is employees’ private information. The assumption is reasonable.

In the model by Hirsch et al. (2017), workers choose their workplace presence depending on their disutility associated with workplace attendance; this is unobservable to their employer. The disutility depends on workers’ health status. The disutility is larger for those who have poor health statuses. To simplify the theoretical analysis, workers are assumed to be risk neutral in the models on employment relations. Workers are usually assumed to be risk averse in the theoretical models because they have only limited opportunities to obtain comprehensive insurance against fluctuations in wages and salaries. Attendance at work is costly for a worker because attendance reduces worker’s utility owing to the lower level of leisure time. Therefore, workers naturally attempt to minimize attendance at work.
However, workers encounter a trade-off. Attendance at work has a counterbalancing effect because attendance decreases the probability that the worker is dismissed from the job. For a worker, dismissal causes a substantial amount of disutility because the search for a new job is costly for the affected worker. Thus, the fear of unemployment (i.e., perception of job insecurity) may increase the prevalence of presenteeism in organizations and firms. There is also earlier evidence according to which sickness absenteeism is at a higher level when the aggregate unemployment rate is lower (Askildsen et al. 2005). Thus, both categories of work-related sickness behaviour may respond significantly to macroeconomic conditions.

To make the model tractable and generate predictions that are consistent with the stylized empirical facts of presenteeism, there are two types of workers in the model by Hirsch et al. (2017). Some workers have low disutility from the work, and other workers have high disutility. Those workers who work while sick show inefficiently high effort. In contrast, those who are absent show inefficiently low effort. Because of the asymmetric information between employers and employees, employers are forced to use only the wage level to financially incentivise workers. Thus, the wage level is the relevant decision variable for a firm.

The presence of asymmetric information implies that employers are forced to set wages to incentivise the “average” worker in the firm. In the absence of asymmetric information, workers would obtain an optimal person-specific wage set by firm. Therefore, in the equilibrium, employers provide overly strong incentives for sick workers and overly weak incentives for healthy workers to attend work. This fundamental feature of the model implies that some sick workers attend work, and other healthy workers are absent from work. Thus, there is a suboptimal allocation and mixture of labour input in the firm that has
a negative impact on a firm’s output and economic performance. Consequently, there are both presenteeism and absenteeism in the model equilibrium; this is the key novel feature of the model compared with earlier theoretical considerations.

The theoretical model presented by Hirsch et al. (2017) has three important implications that can be tested with suitable empirical data. First, the model implies that workers differ notably in terms of their presenteeism behaviour. This implication reflects the underlying fundamental differences in the disutility from workplace attendance among workers. In particular, workers who have poor health statuses and/or encounter significant work-related stressors such as adverse working conditions in their working environments and/or poor management practices that expose them to work-related disamenities should have a significantly higher incidence of presenteeism. Second, workers with longer tenure (i.e., work experience in the current firm) who typically have stronger de facto dismissal protection should have less presenteeism because the probability of dismissal has an effect on the incentives of workers to attend while being sick. A higher probability of dismissal implies that a worker is more likely to attend while sick, all else being equal. The probability of dismissal may be related to an employee’s age. In Sweden, there is an explicit last-in-first-out rule according to which those who have the longest tenure also have the strongest job protection against dismissals. Third, presenteeism should be more relevant for high-skilled workers whose productivity is higher. It is arguably more difficult to find a suitable replacement for a high-skilled worker than for a low-skilled worker who only has a general set of skills. The level of formal educational attainment can be used to measure the skill level of workers in empirical research.
Empirical applications

There are only a very limited number of empirical studies on the issues directly related to presenteeism in economics. As in other disciplines that have examined the determinants and consequences of presenteeism, the existing empirical studies are based on survey data that contain self-reported information on presenteeism and employee characteristics such as educational attainment and self-assessed health. Survey data also contain certain information on firm characteristics such as firm size that may be relevant to better understand the determinants of presenteeism. Surveys are usually cross-sectional; thus, they do not allow researchers to follow the same employers and workers over time. There are empirical studies that strive to quantify the specific costs associated with presenteeism. Importantly, the comprehensive and accurate measurement of costs associated with presenteeism has substantial practical value for organizations and firms that strive to minimize the negative consequences of presenteeism on financial performance.

Most notably, Pauly et al. (2008) provide comprehensive empirical evidence on manager perceptions regarding the costs associated with presenteeism. The objective of the study is to cover various types of jobs that have very different actual content. The data cover 804 middle-level managers from the United States. The sample size is very limited to derive conclusions that would be nationally representative. The data contain detailed questions about the costs of work-related illness and the consequences of working while being sick. Pauly et al. (2008) document that jobs with a particularly high incidence of team production, high requirements for timely delivered output, and high difficulties of substitution for absent workers have significantly higher costs for both absenteeism and presenteeism.
Interestingly, Pauly et al. (2008) conclude that substitution opportunities are a somewhat less important factor for presenteeism than absenteeism. Poor substitution is a particularly relevant feature from the economic perspective because, when substitution between workers is not possible, after he or she returns to work, a (white-collar) worker must accomplish all those tasks that were not done during his or her absence from work. This work significantly increases the incentives to work while being sick but may cause substantial negative externalities to co-workers, as described earlier.

Kigozi et al. (2017) provide a comprehensive summary of the empirical literature that has evaluated the costs that are associated with presenteeism in firms. The systematic review covers 28 studies. One main conclusion of the summary is that the costs associated with presenteeism are only rarely included in economic evaluations.

Kigozi et al. (2017) state that the literature has used two methods to estimate the losses associated with presenteeism. The first method is the direct approach. The method involves relevant respondents to estimate the extra hours that would be needed to compensate for inefficient hours that were caused by working while being sick. The second method that has been used in the literature also uses self-reported information to provide a perceived overall estimate of how much illness has reduced performance at work. This approach has been the most popular method in the literature.

The most important problem with self-reported information is that it may contain a substantial amount of systematic measurement error. For example, personality traits may have an influence on the manner in which respondents answer the questions, and the same characteristics of persons may also affect their work-related sickness behaviour such as
presenteeism. This observation makes it difficult to derive well-grounded conclusions for policy design.

Hirsch et al. (2017) evaluate the empirical implications of their theoretical model using German survey data. The researchers find that presenteeism is typically at the higher level for those workers who have worse health, all else being equal. Health status is measured using self-assessed health. Hirsch et al. (2017) also find that adverse working conditions increase the prevalence of presenteeism. These empirical findings are consistent with the predictions of the theoretical model because the formal model states that a higher (health-related) disutility derived from attendance at work is associated with a higher incidence of presenteeism. However, the association between worker’s tenure and presenteeism is quantitatively weak. Therefore, tenure is not a practically important determinant of presenteeism, at least not according to the German survey data. It is important to note that the effects may be notably different in other institutional contexts. Furthermore, Hirsch et al. (2017) find that the relationship between employees’ educational attainment and presenteeism is also not as clear as the theoretical model implies.

Conclusions

There are two important work-related sickness categories that have been examined in the literature. There is a large body of theoretical and empirical literature in economics regarding the determinants and consequences of absenteeism (Treble and Barmy, 2011). However, economic research has only recently begun to more deeply examine both the determinants and consequences of presenteeism. There are several relevant challenges in future theoretical and empirical studies on presenteeism in economics. The challenges
related to empirical research are also highly relevant for investigations that are conducted in other disciplines such as organization and management studies.

There are important issues related to the existing theoretical models. First, theoretical models that are based on the optimizing behaviour of firms and workers should incorporate more relevant aspects of institutions because they are important determinants of the prevalence of presenteeism. For example, the financial incentives of sickness insurance schemes are directly relevant for the decision whether to work while being sick or not. Second, theoretical models should more fully consider that workers have different preferences towards working while being sick. Preferences may also be related to non-economic determinants of presenteeism such as peer pressure from co-workers to attend work while being sick. Peer pressure may be particularly strong in small, self-managed teams that are currently common in organizations and firms. The integration of economic and non-economic determinants of presenteeism would open new possibilities to understand the underlying motives of presenteeism more deeply. Third, there should be more theoretical modelling that fully considers the empirical fact that absenteeism and presenteeism prevail simultaneously in many organizations and firms. There is also a positive correlation between the prevalence of absenteeism and presenteeism at the individual level (Böckerman and Laukkanen, 2010). Currently, there is only one theoretical framework in economics (Hirsch et al. 2017) that can explain one of the basic stylized facts of the empirical literature (i.e., the simultaneous existence of both absenteeism and presenteeism). Third, theoretical models should consider the negative effects of working while being sick on employees’ health status; this may cause an increase in the amount of absence for sickness in the long run. These long-term effects have been ignored in the
theoretical models, but they are nevertheless important for the complete evaluation of the costs related to presenteeism.

There are also equally important challenges related to the empirical perspective of the research on presenteeism. First, research should be based on nationally representative data that are available for many countries. Most of the empirical literature focuses on a narrow set of industries and firms. A key problem is that both the motives and consequences of presenteeism are most likely significantly heterogeneous across organizations. Thus, the specific organizations that have garnered researchers’ focus may be those where the effects are anticipated and/or are according to prior theoretical considerations. This supposition makes it particularly difficult to extrapolate from these results to the population as a whole. This hindrance is an important issue because there is earlier evidence supporting the notion that the determinants and consequences of presenteeism are heterogenous across industries and firms. Nationally representative data are also needed to derive policy conclusions that can be implemented in a broad range of firms. Thus, this apparent heterogeneity calls for more studies. It would also be particularly important to consider this heterogeneity in the design of management policies to reduce the prevalence of both absenteeism and presenteeism because different groups of workers clearly respond differently to economic and non-economic incentives. Therefore, the average effects on which the literature has greatly focused on provide us with an incomplete picture of the determinants of absenteeism and presenteeism compared to a more nuanced setting that would allow for the heterogeneity of the effects.

Second, workers are clearly not randomly assigned to firms. This statement implies that the correlations that are calculated using observational data (e.g., the correlations between
employee characteristics such as educational attainment and presenteeism) do not reveal causal effects about the underlying determinants of presenteeism. Thus, the most important difficulty in interpreting the cross-sectional correlations is establishing whether the relationships between background characteristics and presenteeism are causal or not. Therefore, the sorting of workers into workplaces and job tasks within firms cannot be ignored. In particular, if jobs that have a higher incidence of various stressors also attract more able and healthier employees, or those who are mentally and physically more resilient towards stressors, the cross-sectional correlations between individual characteristics and the measures of work-related sickness behaviour will be biased, at least to a degree. The size of this bias is not known. One tractable approach with linked survey and register data is to use information on employees’ wage and work histories. Therefore, to alleviate this problem, it is useful to condition based on employees’ work and earnings histories, which are most likely highly correlated with unobserved worker traits, thus reducing the potential for omitted variables bias. A key challenge for empirical studies is that the use of work and earnings histories requires the use of linked survey and register data in which work and earnings histories are derived from comprehensive register-based data. Linking survey data to register-based information on employees’ work and wage histories is possible in Nordic countries by utilizing person-specific identification codes (Böckerman et al. 2012; 2013).

Third, most of the empirical research on presenteeism uses cross-sectional data. The use of panel data would allow researchers to fully control for person-specific fixed effects such as time-invariant personality traits that are potentially important determinants of both absenteeism and presenteeism. Fully considering personality traits and other time-invariant characteristics would allow researchers to more precisely identify the independent effects of the economic determinants of presenteeism. This identification would provide valuable
information to design policies to reduce presenteeism and the negative externalities that it causes in organizations and firms.

Fourth, there are important gaps in the literature regarding more specific issues. One neglected aspect of research has been the effect of new management initiatives on work-related sickness. High involvement management practices have become an increasingly popular element of contemporary human resource management. These practices contain several different aspects such as self-managed teams, problem-solving groups, information sharing, incentive pay and supportive practices such as employer-provided training and associated recruitment methods. There is also most likely a substantial amount of heterogeneity in the effects of new management initiatives on presenteeism by gender, age groups and type of employer. Future research should focus more on the potential role of high involvement management practices in the determination of presenteeism.

Fifth, economic research on the determinants and consequences of presenteeism would benefit greatly from the use of linked survey and register data. The measures of presenteeism and employee characteristics are nearly always based solely on survey information. However, by linking comprehensive register-based information from employees and firms, it is possible to broaden the potential determinants of presenteeism and provide additional valuable lessons for policy makers.
References


