Manchester Metropolitan University

e-space

Manchester Metropolitan University's Research Repository

Please cite the Published Version

Skagen, Kristian and Collins, Alison M (2016) The consequences of sickness presenteeism on health and wellbeing over time: A systematic review. Social Science & Medicine, 161. pp. 169-177. ISSN 0277-9536

DOI: https://doi.org/10.1016/j.socscimed.2016.06.005

Publisher: Elsevier

Version: Accepted Version

Downloaded from: https://e-space.mmu.ac.uk/621650/

Usage rights: Creative Commons: Attribution-Noncommercial-No Derivative Works 4.0

Additional Information: This is an Author Accepted Manuscript of a paper accepted for publication in Social Science & Medicine, published by and copyright Elsevier.

Enquiries:
If you have questions about this document, contact openresearch@mmu.ac.uk. Please include the URL of the record in e-space. If you believe that your, or a third party's rights have been compromised through this document please see our Take Down policy (available from https://www.mmu.ac.uk/library/using-the-library/policies-and-guidelines)

Penultimate version. If citing, please refer instead to the published version in Social Science and Medicine, 2016, Vol 161, 169-177 doi:10.1016/j.socscimed.2016.06.005

The Consequences of Sickness Presenteeism on Health and Wellbeing over Time: A Systematic Review

Kristian Skagen[1], Alison M Collins[2]

[1]Department of Business and Economics, Centre of Health Economic Research, University of Southern Denmark, Campusvej 55, 5230 Odense, Denmark

[2]Centre for Organizational Health and Wellbeing, Lancaster University, LA1 4YG, United Kingdom

Abstract:

Rationale

The association between sickness presenteeism, defined as going to work despite illness, and different health outcomes is increasingly being recognized as a significant and relevant area of research. However, the long term effects on future employee health are less well understood, and to date there has been no review of the empirical evidence. The aim of this systematic review was to present a summary of the sickness presenteeism evidence so far in relation to health and wellbeing over time.

Methods

Eight databases were searched for longitudinal studies that investigated the consequences of workplace sickness presenteeism, had a baseline and at least one follow-up point, and included at least one specific measure of sickness presenteeism. Of the 453 papers identified, 12 studies met the eligibility criteria and were included in the review.

Findings

We adopted a thematic approach to the analysis because of the heterogeneous nature of the sickness presenteeism research. The majority of studies found that sickness presenteeism at baseline is a risk factor for future sickness absence and decreased self-rated health. However, our findings highlight that a consensus has not yet been reached in terms of physical and mental health. This is because the longitudinal studies included in this review adopt a wide variety of approaches including the definition of sickness presenteeism, recall periods, measures used and different statistical approaches which is problematic if this research area is to advance. Future research directions are discussed.

Key words: sickness presenteeism, health outcomes, physical health, mental health, self-rated health, well-being, longitudinal studies, sickness absenteeism, systematic review

Introduction

Sickness presenteeism (SP) has been receiving a growing amount of attention among researchers because of its potential impact upon both the organization and the employee. As Johns (2010:521) points out, the concept of presenteeism has suffered from "definitional creep", as the term has been defined in a number of ways over time. Recently the focus on the term SP has centred on two definitions: One emphasises the economic consequences of employees absence, by for example, measuring productivity losses as a result of chronic health conditions such as arthritis, migraine, allergies, depression/anxiety and cancer and has primarily been researched from a North American context (see for example Schultz and Edington 2007, for a review). A second emphasis, and the focus of this review, is the health consequences to employees reporting SP, defined by Aronsson et al., (2000:503) as "people, despite complaints and ill health that should prompt rest and absence from work, still turning up at their jobs". This definition of SP incorporates health in general terms and does not focus on specific chronic conditions; it concerns individuals who go to work despite being ill and the work-related and personal factors that influence SP. The prevalence of SP, defined as going to work while being ill one or more times during a pre-defined period, ranges from 47% in a Swedish police cohort (Leineweber et al., 2011) to 73% in a Danish workforce sample (Hansen and Andersen, 2008) using a single item question and a recall period of the previous 12 months. Thus, it appears that SP is a common organizational behaviour.

Although, as Johns (2010) points out research around SP has been largely atheoretical, Aronsson and Gustafsson (2005) proposed a model for future research into SP. Aronsson and Gustafsson (2005) suggest that the decision to attend work when ill or take SA, is influenced by attendance demands which can be personal and/or work related. Personal factors include having a conservative attitude to taking sick leave (Hansen and Andersen, 2008), boundarylessness (the ability to say no), and financial constraints (Aronsson and Gustafsson, 2005). Work demands include time pressures,

the worker's replaceability, (Aronsson and Gustafsson, 2005) and workload (e.g. Biron et al., 2006). Overall, there has been some support for the first part of Aronsson and Gustafsson's (2005) model which concentrates on the demands which lead to SP (e.g. Hansen and Andersen, 2008; Claes, 2011). However, there has been less research into the second half of their model which focusses on whether SP affects an individual's future health.

Aronsson et al., (2000) suggest that SA can promote health by allowing time for physical and psychological recovery. Kristensen (1991) found that slaughterhouse workers used SA as a way of coping with work conditions. Indeed, taking short spells of time off work may allow individuals who are ill or stressed to recover, so that more serious conditions are avoided (Kristensen, 1991). The ability to recover from work appears to be important as Kivimaki et al (2005) found that employees who rarely recovered from work during free weekends had an increased risk of death through cardiovascular disease. In contrast to SA, SP has been associated with more negative outcomes such as productivity loss, aggravating existing conditions (Johns, 2010) and negatively affecting the health of colleagues (Baker-McClearn et al., 2010). Studies based on cross-sectional data suggest that higher levels of SP is associated with poorer self-rated health (Aronsson and Gustafsson 2005; Leinweber et al., 2011), higher levels of psychological distress and psychosomatic complaints (Biron et al., 2006), reduced physical and mental health and increased exhaustion (Lu et al., 2014)

The relatively few longitudinal studies that examine the impact of SP on health outcomes over time indicate that SP at baseline may have future negative health and wellbeing consequences for an individual. For example, SP has been found to be an independent predictor of future poor self-rated health (e.g Bergström et al., 2009a; Gustafsson and Marklund, 2011; Dellve et al., 2011). This may have significant implications when we consider that self-rated health has been found to be a good

predictor of mortality (Fayers and Springer, 2002). Indeed, Kivimaki et al., (2005) found that males with poor health who did not take SA over a two-year period had twice the risk of cardiovascular disease than men who took moderate (0-14 days) SA. Thus, this is potentially an important area of research. However, a consensus is still to be reached in terms of the consequences of SP over time. This is because researchers adopt diverse research designs, and while studies include similar health and wellbeing outcomes, such as physical and mental health, how they are measured and the data analysed vary greatly across studies. The aim of this systematic review, therefore, is to explore the impact of SP on future health and wellbeing. By focussing on SP our review differs from Schultz and Edington (2007) which explored the links between health and productivity presenteeism. In this paper, we take a wide view of wellbeing that takes into consideration the whole person in order to encompass a range of outcomes of SP that have been included in the selected papers. It should be noted that studies concentrated upon negative health outcomes, and did not consider the positive outcomes that workers with, for example, chronic health conditions may experience by remaining in work.

Methods

This review followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses PRISMA guidelines (Liberati et al., 2009). Prior to carrying out the systematic literature search, the inclusion and exclusion criteria were specified and documented in a research protocol. The criteria for inclusion were that studies adopted a prospective study design with at least one follow-up after baseline (initial measurement), included an explicit measure of SP and had been published in a peer-reviewed journal. Studies that focussed on economic evaluations or employees with chronic diseases or adopted a retrospective approach were excluded.

Search Strategy

The literature search was conducted in the following electronic databases with no limits applied for language, as non-English written articles were translated: PubMed Medline (1966-present), Web of Science (1945-present), EconLit (1968-present), Academic Search Complete (1970-present), EMBASE (1970-present), CINAHL (Cumulative Index to Nursing and Applied Health) (1961-present), PsychINFO (1967-present) and Directory of Open Access Journals (2003-present). The last search was conducted on the 28th of January 2015. Furthermore we scanned reference lists to make the search as wide as possible. The following search string was applied to search all the databases: Presenteeism OR "sickness attendance" OR “SP” AND “cohort OR prospective OR follow-up OR panel OR longitudinal”. If there was any doubt whether an article should be excluded or not it was included to the following stage. Abstracts for all the included articles were retrieved, and each abstract was screened independently by both authors. The reason for exclusion was recorded for all excluded articles. A list of articles was drawn up and compared by both authors, and any disagreements were resolved by consensus.

Data extraction and quality assessment

We developed a data extraction tool that took into consideration the review question (Khan et al., 2001). The developed tool was pilot tested on two articles leading to minor corrections to the extraction tool. The two review authors extracted the data from included studies and the extractions were compared and documented. Due to the heterogeneous nature of SP studies in terms of SP definitions, differing methods, as well as different outcome measurements, we undertook a thematic approach to the review. We adopted an interpretive approach in an attempt to broaden our understanding of SP by identifying key emerging themes. Using quality scoring scales, which generate a numerical summary score and weight one item over another, to assess study quality is no longer encouraged (Higgins, 2011). We therefore assessed each paper for quality by identifying the

strengths and limitations of the study design and methods that limit bias and increase internal validity (Crowe and Sheppard, 2011; Higgins, 2011; Sanderson et al., 2007). The dimensions we considered were:

- Length of time between baseline and follow-up(s)
- The sampling method: to examine whether the participants in the study are representative of the target population
- The appropriateness of the sample size
- Control of variables: including the appropriateness of control variables, and omitted variables
- Measurement of health and wellbeing outcome variables (including use of validated instruments)
- How attrition from the studies was managed

We applied the above dimensions to the included papers to identify strengths and limitations, summarised in Table 1. We concluded that the quality of the studies did not differ greatly and therefore treated all included studies equally.

Results

A total of 453 papers were found through the literature search (Figure 1) and duplicates were removed. The remaining abstracts were reviewed by title, and those which did not meet the inclusion criteria were excluded (N=356). The remaining 60 papers were assessed by their abstracts and 46 were excluded. A further two articles were excluded after we had read the full text. A total of 12 papers were included for the final review which are predominantly European (primarily Scandinavian), with two coming from Taiwan (Lu et al., 2013 and 2014). All reviewed papers were quantitative and adopted a survey-based design. Ten studies drew upon survey data from both

public and private sector workplaces, incorporating different occupations. Two studies (Dellve et al., 2011; Demerouti et al., 2009) focus solely on the public sector. The sample size varied across the studies. Gustafsson and Marklund (2011and 2013) drew upon the same sample population, as did Bergström et al., (2009a and 2009b). The two papers by Lu et al., (2013 and 2014) also used the same population; however, Lu et al., (2014) additionally included 100 Chinese respondents in the sample. One of the studies reviewed considered health outcomes in terms of sickness levels, and adopted both SP and SA as predictor variables (Gustafsson and Marklund, 2013).

The organizational and individual factors which influence SP were not included in the survey design of the reviewed studies, so the underlying influences or reasons why people go to work despite being ill were not considered. All papers included self-reported SP and measured the impact upon associated self-reported health outcomes over time (see Table 1). Inclusion of health outcomes in this review were data-driven by the included studies and comprised self-rated health, physical health, mental health, the employees' assessment of their ability to work, and sickness absence.

Insert Figure 1 here

The health consequences of presenteeism over time

More detailed information on the results of the included studies is available as supplementary data.

Self-rated health

SP at baseline was found to be an independent predictor of future poor self-rated health (SRH) at 12 months (Gustafsson and Marklund, 2011) and 24 months follow-up (Dellve et al., 2011). Bergström et al., (2009a) reported a similar outcome in a two-wave study at 18 and 36 months follow-up.

Furthermore, Bergström et al., (2009a) found that the impact of SP was only significant among public sector employees with good/excellent health at baseline, while SP significantly impacted upon the future health of private sector employees, regardless of their health status at baseline. The reason for this difference is unclear. It may be due to differences between the two sectors' populations or because the number of respondents who dropped out of the study was greater among those with fair/poor health and high SP (Bergström et al., 2009a). The frequency of reported SP appears to be important with a dose-response relationship, that is, as the number of SP days increases so does the risk of future poor SRH (Bergström et al., 2009a; Gustafsson and Marklund, 2011). However, while Taloyan et al., (2012) found that seven or more days of SP at baseline increased the risk of future decreased SRH; this association was attenuated to non-significance when emotional exhaustion at baseline was controlled for. When SA and SP were both adopted as predictor variables, the risks for future SRH were higher for those individuals with greater levels of sickness, regardless of their exhibited attendance behaviour (Gustafsson and Marklund, 2013).

Physical health

The studies measured physical health in different ways. For example, Gustafsson and Marklund (2011 and 2013) measured physical complaints (e.g., pain in upper/lower back, shoulder hips and wrists) and found both SP and SA were predictors of future physical health complaints. Further analysis revealed that respondents with higher levels of sickness, reporting both high SP and SA had a higher risk for physical complaints a year later (Gustafsson and Marklund, 2013). In contrast, using a six-item Physical Health scale, Lu et al., (2013) found an association between SP at baseline and physical health at T1, two months later. However, no association was found at a three-month follow up (Lu et al., 2014). Thus, there is limited evidence to suggest that SP may impact upon future physical health.

Mental health

A variety of mental health measures were adopted in the reviewed studies. For example, SP of eight or more days at baseline was associated with a 2.45-fold increased risk of depression at follow-up amongst non-depressed participants at baseline (Conway et al., 2014). Lu et al., (2013), using a 12-item mental health scale, found an association between SP at baseline and mental health at T1 two months later. Gustafsson and Marklund (2011) utilising a 10-item mental wellbeing scale, which incorporated one depression item, one anxiety item and other items around positive wellbeing/energy found that SP impacted upon mental wellbeing at T2, twelve months later. When looking at SP and SA in combination Gustafsson and Marklund (2013) found that the odds ratio for poor mental wellbeing at T2 was raised for participants with higher levels of sickness, reporting both SP and SA, and further raised for respondents with high SP (>2 times)/ low SA (0-1 time) .

Demerouti et al., (2009) measured two dimensions of burnout: Emotional exhaustion and depersonalisation. They found a reciprocal relationship between emotional exhaustion and SP, in that exhaustion at baseline led to SP at T1 (at 12 months), which in turn led to increased exhaustion at T2 (six months later). In addition, they found that SP led to increased depersonalisation over time. Dellve et al., (2011) also found that SP led to burnout two years later. Lu et al., (2013) found that SP was related to exhaustion at T1, two months later. However, Lu et al., (2014) found that SP did not predict mental health at three months follow-up. In summary, there is mixed evidence relating to SP and mental health, although it appears that working while being ill may increase the risk of depression.

Work ability

The work ability index (WAI) is a self-assessment tool in which respondents provide information about their general state of health and estimate their ability to work. Gustafsson and Marklund, (2011 and 2013) and Dellve et al., (2011) found that SP of two or more days was found to be a

predictor of future poor work ability. However, Gustafsson and Marklund (2011) found the odds ratio for future low workability was higher amongst respondents reporting SP five days or more than those reporting two to five days, suggesting a dose response effect. In addition, Gustafsson and Marklund, (2011) found that higher levels of SA (3 weeks or more) had a similar negative impact upon participants future poor work ability. Thus, there is limited evidence to suggest that both SP and SA negatively impact on future work ability.

Sickness absence

Overall, the findings from the reviewed studies that included SA as an outcome variable suggest that two or more occurrences of SP at baseline significantly increase the risk of future short and long term SA. For example, two or more episodes of SP increased the risk of future SA of at least two weeks' duration (Hansen and Andersen, 2009). In addition, respondents with six or more occurrences of SP at baseline increased the risk of long-term SA of at least two weeks by 53% and increased their risk of SA for two months or more by 74% at 18 months follow-up (Hansen and Andersen, 2009). SP of five or more days at baseline was significantly associated with SA of one week or more (Gustafsson and Marklund, 2011) as well as short periods of SA (between 1 and 3 days) and long periods (at least 15 days) of SA at 12 months follow-up (Janssens et al., 2013). Bergström et al., (2009b) found that SP of five or more days increased the risk of future SA of 30 or more days for private sector employees at 18 months and 36 months follow-up. However, while Taloyan et al., (2012) initially found that seven or more days of SP at baseline increased the risk of having more than seven days of SA by 46% after two years, this result was attenuated after controlling for emotional exhaustion. They suggested that the health consequences of SP over time may be due to mental health.

Insert table 1 here

Methodological differences

Variation across studies

While the studies measured similar health and wellbeing outcomes, they adopted diverse research designs. Firstly, all the reviewed studies adopted a single question to measure SRH, with slight variations in phrasing. The majority of studies also relied on a single question to measure SP, which does not include the circumstances surrounding the decision to go to work despite being ill. Only Lu et al., (2013 and 2014) incorporated a wider assessment of SP to include avoidance motives (concerns about being punished or being perceived negatively by colleagues or supervisors) and approach motives (such as being loyal to their job and colleagues) underlying the decision to go to work while ill. In addition, the subjective nature of SP means that studies have to rely on self-reporting. A similar recall period was adopted by 10 of the studies, requiring participants to remember their attendance behaviour over a 12 month period. Only two of the studies by Lu et al., (2013 and 2014) adopted a shorter recall period of six months. The authors did not provide a reason for this decision, although it is likely that a shorter recall period was chosen to promote a more accurate memory recall.

Even though many of the studies measured the same health outcomes, the variation in measures used means that results are difficult to compare across studies. For example, the Work Ability Index was adopted in two studies, however, it was utilised very differently: Gustafsson and Marklund (2011 and 2013) adopted the five-item WAI scale and scored work ability as either high or low, while Dellve et al., (2011) adopted the 10-item WAI scale and scored work ability as poor, moderate, good or excellent. In addition, the length of time between data collection waves varied

from two to 36 months (see Table 1). This can make it difficult to compare across studies, however, studies that do adopt different times between follow-ups can add to our knowledge of the consequences of SP. As the study by Lu et al., (2013) indicates, SP may have short term consequences as well as long term.

A key feature of SP research is that a large percentage of respondents report no SP or SA during the previous year, or report just one instance. For example, around 28% of respondents may not report any SP within a specific recall period (e.g. Bergström et al., 2009a and 2009b; Gustafsson and Marklund, 2011) whilst those reporting just one instance may increase this 'healthy' group to almost half of the sample (e.g. Janssens et al., 2013). Thus, while ten of the review studies used a similarly worded question to ascertain the occurrence of SP, varying data categories were adopted, presumably to take into consideration the low SP reported by respondents. For example, Gustafsson and Marklund (2011 and 2013) adopted a five response option in the questionnaire which was then analysed in terms of four categories never/once/2-5 times/>5 times. In contrast, Demerouti et al., (2009) adopted two response options of no and yes (thereby not measuring SP frequency or duration) in order to reduce recall response errors. How researchers distribute the SP data across categories has implications for how the findings are interpreted in terms of the number of SP occurrences required to influence health over time. In addition, it is difficult to compare across findings.

Three different approaches towards inclusion of control variables in the analysis were found (see Table 1). One approach to the statistical analysis of control variables was stepwise selection, either forward and/or backward, this entails a data-driven method, where each control variable was taken out of the subsequent analysis if it changed the result less than 10% (Bergström et al., 2009a and 2009b). Two other studies used a slightly different approach where they included those control

variables that were found to be significant when testing the model (Janssens et al., 2013; Taloyan et al., 2012). A second approach was to use blocks of control variables to build different models, entering one block of control variables at a time (e.g. Conway et al., 2014; Gustafsson and Marklund, 2011 and 2013), relying on a data driven approach rather than prior theoretical considerations. A third method was a more theory guided approach and included a wide number of control variables found to be associated with SA in the literature. As can be seen from Table 1, although Hansen and Andersen (2009) used a theory guided approach, and Bergström et al., (2009b) used a data driven approach to investigate whether SP at baseline increases the risk of SA at T2, they arrived at similar results.

Discussion and future research agenda

The aims of this review were to to explore the impact of sickness presenteeism on future health and wellbeing. It was important for a systematic review to be undertaken in this research area, as it provides evidence of the extent of variation within the SP literature in terms of research design, the measurement and analysis of SP, and health and wellbeing outcomes that make it difficult to compare results across studies. Thus, there is a clear need for further qualitative and quantitative research to further our understanding of this phenomenon. In the following sections we discuss the findings from the reviewed studies and suggest a research agenda.

Overall, the reviewed studies support a link between SP at baseline and future poor self-rated health and SA. With mental health there is some evidence to suggest that working while ill may increase the risk of depression over a two-year period. It appears that those individuals presenting with higher levels of sickness, regardless of their exhibited attendance behaviour, have a higher risk for future self-rated health. Aronsson and Gustafsson (2005) suggest that ill-health and a loss of capacity are the main determinants of both SP and SA; however, ultimately the decision to take sick

leave or go to work despite being ill is dependent upon personal and work related demands. As Aronsson and Gustaffson (2005) and Johns (2010) point out there are consequences to the individual for both attendance behaviours. For example, while taking sickness absence may have recuperative benefits there may be organisational consequences such as dismissal or a loss of pay. Alternatively, SP may aggravate a health condition which may subsequently lead to SA. Thus, those individuals who present with both SP and SA during the preceding year may have been subject to two different sets of consequences. This may impact upon the individual to a greater extent than the consequences of a single episode of SP or SA. Furthermore, such consequences to the individual could be cumulative and impact upon future heath, attendance, and organizational relationships (Johns, 2010).

The majority of studies rely on a single item to measure SP, which does not include the circumstances surrounding the decision to go to work despite being ill. Only Lu et al., (2013 and 2014) incorporated a wider assessment of SP. We therefore suggest that more detailed and, if possible, objective SP measures are incorporated in future studies. For example, Johns (2010) highlights how a diversity of methods increased understanding of absenteeism and similar methods could be adopted in sickness presenteeism research. For example, Hackett et al., (1989) used daily diaries over four to five months to measure events related to decisions related to taking SA which, similarly to SP, is a relatively infrequent event. The advantage of daily diaries is that they collect data in real time, and therefore recall around the decision to go to work whilst ill would not be affected by poor memory. In addition, daily diaries would potentially give more detailed data which would allow researchers to explore the underlying reasons for SP, the types of illness that are associated with SP as well as personal and organizational factors that influence the decision processes involved in deciding whether to go to work whilst ill

People with chronic conditions will potentially have more SP events than healthy individuals. Therefore, an event-contingent diary approach, whereby data recording would occur whenever a decision to either take sick leave or SP, could be useful in exploring the positive, or negative, health outcomes associated with an attendance decision. While, a review by Schultz and Edington (2007) found that chronic health conditions are associated with productivity losses at the workplace, the long-term health consequences of SP for people with such conditions are less understood. Indeed, none of the longitudinal studies overtly considered employees with chronic health conditions, in that the studies did not measure chronic conditions, although they may have been included in the population sample. Future research, which differentiates between healthy employees and those with chronic conditions, is warranted in order to further our understanding of the consequences of SP for different types of illnesses and health conditions, as potentially, different health conditions may lead to different outcomes over time.

The vast majority of studies included a follow-up period of at least a year. However, one study Lu et al., (2013) suggested that the consequences of SP can occur after just two months. Further exploration of both the short and long-term association of SP with respect to health outcomes is warranted. We therefore suggest that additional follow-up periods or multiple follow-up periods are incorporated in studies, for example at least every three months, in order to ascertain both the short-term and long-term consequences of SP.

The reviewed studies categorised SP into wide and differing response categories. Using pre-defined categories in the questionnaire can make it easier for respondents to estimate or remember the number of days they experienced SP; however, it limits the availability of possible statistical analysis, and can entail a loss of information. For example, when the response category for SP days ranges from 2 to 30, respondents who report a small number of SP occurrences are likely to significantly differ from those at the other end of the range with regard to health status and SA. This

makes it difficult, not only to compare across studies, but also to ascertain how many days of SP result in negative health consequences. Using count data gives the researcher the possibility to estimate the marginal contribution of experiencing one more day of SP in relation to a certain outcome. In order to establish an understanding of the number of SP occurrences required to impact upon health, we suggest that researchers request the actual number of SP days experienced by the participant as well as the underlying reason for SP. This could be done by including a question about the nature of the illness causing SP in a survey.

While the reviewed studies all attempted to minimise bias, they did not fully control for all forms of bias. For example, in two-thirds of the studies there was a risk of attrition bias because they either did not outline any analysis of attrition in their study population, or just examined the differences between responders at baseline and follow-up with characteristics such as age or gender. In the case of self-rated health, the reviewed studies adopted slight variations in phrasing. However, this is unlikely to be an issue, as previous research suggests that even when versions of the single item self-rated health question vary greatly, responses tend to be similar despite such variations, and the measure has been found to be a reliable predictor of mortality (Fayers and Sprangers, 2002) and SA (Roelen et al., 2014).

There was a potential for selection bias amongst the reviewed studies due to the inclusion of small and selective samples, such as concentrating on workers from a particular population such as nursing staff (Demerouti et al., 2009) or participants attending university classes (Lu et al., 2013 and 2014). In addition, some studies also excluded respondents who were on sick leave at baseline or follow-up, or had reported extensive sick leave in the previous 12 months. Thus, it can be argued that included participants were healthier than the overall sample. These different approaches could explain some of the variation in the findings and could potentially bias the results of the studies if

non-respondents differ from those included in terms of health, SA and other measured outcome variables. For example, Bergström et al., (2009a) found a higher non-response rate among employees reporting greater levels of SP, especially for those respondents who reported fair/poor health.

There was a potential for recall bias in ten of the reviewed studies (Lu et al., 2013 and 2014 are the exceptions). Although it can be argued that using a 12 month time period does help to eliminate the potential problem of seasonal variations in the data due to, for example, flu seasons during the winter time, and allergies during the spring and summer time, studies do risk recall bias. Comparisons can be made with SA research which uses similar recall time periods, but sometimes can utilise data from other sources such as organizational sickness records. However, studies that have compared participant recall with such records have found mixed results. For example, Voss et al., (2008) found that participants tend to under-report the number of sick leave days taken over a year compared to company data, although overall they found good agreement between both sources of data, especially when reporting no, 1-7 and over 28 sick days. In contrast, Severens et al., (2000) and van Poppel et al., (2002) found the length of the recall period for retrospective SA to be limited and suggested a recall period of no more than two months. As Johns (2010) points out the temporal stability for SP recall still needs to be established, and therefore greater attention to the "optimal" recall period is required. Until more objective measures of SP can be developed, we suggest that future research adopts shorter, more frequent recall periods that would help to establish an optimal recall period, and thereby minimise the risk of recall bias.

The reviewed studies took three different approaches to how control variables were included in the analysis (see Table 1), all of which are widely used in research. While, it can be argued that a data driven approach is the best solution when there is no prior literature in the research, it may be

difficult to ascertain how many control variables were adopted from the original dataset. Adopting a theory driven approach allows researchers to control for the determinants of SP in their analyses. However, what is controlled for is important. For example, nearly half of studies omitted to control for one or more variables that could potentially be important confounders of health outcomes over time, such as controlling for SA and self-rated health at baseline. As Becker (2005) argues, when selecting control variables researchers should provide a reason why the variable was included otherwise it could lead to bias, and include evidence from the literature that supports the inclusion. In addition, Becker (2005) recommends that researchers should be aware of including impotent variables since this could reduce power. In future studies, researchers should consider taking into account the control variables already identified in previous research and consider what further potential control variables should be considered. Where possible, questionnaires should be specifically designed to include potential control variables relating to SP, because as Becker (2005:286) points out "[t]he results of a study often depend on what control variables are included in the analyses."

The relatively long time between follow-ups used in the majority of the studies makes it possible for the respondents to experience an illnesses or health condition that may not have been caused by SP. For example, a participant may have reported several days of SP at baseline due to having the flu. However, at follow-up they may have developed another illness which is not related to SP at baseline, such as cancer. We argue that it is important to control for initial health, and chronic conditions, at baseline since this is an important determinant of future health associated with both SP and SA. In addition, it is also important to control for changes in health, for example, those who have developed chronic conditions between baseline and T1. In those cases where data is available on health at baseline and at follow-up it would improve the quality of the studies by controlling for the difference score in health (SRH at T1 minus SRH at baseline). This approach was adopted by

Lu et al., (2013) who used the difference score for SP effectively making each participant act as their own control (Allison, 2005), thereby making it possible to control for the health change that might have happened over time.

The cultural setting in which the studies take place also need to be considered. The majority of countries included in this review are European; however, there are differences in the welfare systems that could explain variations in the results. For example, as Claes (2011) points out, the high social expenditure for welfare systems, financed through taxation in countries such as Denmark and Sweden, may inhibit SP. In contrast, the welfare systems in countries such as the United Kingdom may encourage SP. Thus, differences within countries' welfare systems as well as cultural norms are likely to influence the prevalence of SP and SA. Further exploration of SP across different cultures is warranted, especially as the majority of studies looking at the long term consequences of SP are Scandinavian.

In addition, the organizational culture , for example SA and return to work policies, in which SP occurs is not explored with a single measure of SP, despite previous cross-sectional research finding this to be important (e.g. Baker-McClearn et al., 2010). For example, the availability of flexible working arrangements in this decision process may also be important and should be explored, as the option to work from home and thereby avoid taking sick leave may be a strategy that employees adopt instead of going to work whilst ill. Currently, the influence of organizational culture upon SP decisions and subsequent future health is unknown, and further empirical work is needed.

The reviewed studies did not explore which health conditions are deemed to be serious enough to meet the "sick" criteria of the single SP measure by respondents, although a study by Collins and Cartwright (2012) suggested that participants did differentiate between the types of health

conditions they would take sick leave for. Again, the use of diary studies to record the type of health condition or illness and the number of days per SP episode would help to establish this. The type of illness is likely to be key when considering future health problems, because intuitively it is difficult to conceive that an individual reporting several days of SP with a cold will have the same risks as an individual with, for example, cardiovascular disease. Future studies could differentiate between illnesses by distinguishing between SP caused by chronic conditions and acute illnesses, such as the flu or a cold. As Bergström et al., (2009a) point out, it needs to be ascertained whether SP is an independent risk factor for future negative health outcomes irrespective of the health condition.

Johns (2010) has previously argued that the interplay between SA and SP should be explored in terms of employee health and well-being. The majority of studies reviewed have considered both SA and SP as predictor variables. Only two studies (Gustafsson and Marklund, 2013, Dellve et al., 2011) have considered these associations in combination and adopted SA and SP as outcome variables in order to explore the health consequences for participants with higher levels of sickness. Gustafsson and Marklund (2013) suggest that a combination of SA and SP is likely to be a more accurate indicator of an individual's future health and wellbeing. Intuitively, this seems probable, as participants with both high SA and high SP are likely to have greater negative associations upon their health than those with only SA or SP. It seems that individuals who exhibit both SP and SA somehow compromise their recovery and this presents in future impaired health. Furthermore, it may be that any recuperative effect of SA is cancelled out by SP. Certainly, the Gustafsson and Marklund (2013) study suggests that the combinations of SA and SP may impact upon physical and mental health in different ways and we suggest that further exploration of how this may occur is warranted. In addition, taking into consideration both SA and SP gives a more rounded picture of the attendance patterns of respondents. As outlined above, a large percentage of respondents report no SP or SA over the recall period. However, respondents who report no SP may either be healthy

or have taken sick leave during that period. By the same token a respondent who reports no SA may either be healthy or has gone to work whilst ill. These dynamics are likely to be influenced by both organizational and individual factors and needs to be explored further.

Limitations

We aimed to incorporate detailed information as outlined in the methods section from as many of the 12 studies as possible; however, not all articles contained the necessary details and that may be a threat to the validity of the present review. In addition, although we conducted a systematic review, the majority of studies showed an association between SP at baseline and an increased risk of a negative future outcome and thus there may some publication bias. However, we have included two studies (Taloyan et al., 2012; Lu et al., 2014) which did not find an association between SP and future health. We did not assess the reviewed papers using a scoring system, or rate them in terms of methodological quality. We do, however, consider and discuss the strengths and limitations of each paper in terms of the study design and methods that limit bias.

Conclusion

We included 12 longitudinal studies from five countries that examined the consequences of SP on health and wellbeing over time. Although the research is diverse in terms of measurement, methods and risk of bias we conclude that SP is a risk factor for future sickness absence and decreased self-rated health. However, our findings highlight that a consensus is yet to be reached in terms of physical and mental health, and workability outcomes. The extent to which the included studies minimised bias varied. The risk of omitted variable and selection bias is especially of concern in several studies. There are relatively few longitudinal studies within SP research, but they are valuable and give us new knowledge compared to cross-sectional studies. We suggest further

longitudinal research should be undertaken to further our understanding of this phenomenon, which examine the cultural, organizational and personal context in which decisions to go to work despite illness are taken.

References

Allison, P.D. (2005). *Fixed effects regression methods for longitudinal data using SAS*: SAS Institute.

Aronsson, G., and Gustafsson, K. (2005). Sickness Presenteeism: Prevalence, Attendance-Pressure Factors, and an Outline of a Model for Research. *Journal of Occupational and Environmental Medicine*, 47, 958-966. doi 10.1097/01.jom.0000177219.75677.17

Aronsson, G., Gustafsson, K., and Dallner, M. (2000). Sick but yet at work. An empirical study of sickness presenteeism. *Journal of Epidemiology and Community Health*, 54, 502-509. doi 10.1136/jech.54.7.502

Baker-McClearn, D., Greasley, K., Dale, J., and Griffith, F. (2010). Absence management and presenteeism: the pressures on employees to attend work and the impact of attendance on performance. *Human Resource Management Journal*. doi 10.1111/j.1748-8583.2009.00118.x

Becker, T.E. (2005). Potential problems in the statistical control of variables in organizational research: A qualitative analysis with recommendations. *Organizational Research Methods*, 8, 274-289. doi 10.1177/1094428105278021

Bergström, G., Bodin, L., Aronsson, G., Hagberg, J., Lindh, T., and Josephson, M. (2009a). Does sickness presenteeism have an impact on future general health? *International Archives of Occupational and Environmental Health*, 82, 1179-1190. doi 10.1007/s00420-009-0433-6

Bergström, G., Bodin, L., Hagberg, J., Aronsson, G., and Josephson, M. (2009b). Sickness presenteeism today, sickness absenteeism tomorrow? A prospective study on sickness presenteeism and future sickness absenteeism. *J Occup Environ Med*, 51, 629-638. doi 10.1097/JOM.0b013e3181a8281b

Biron, C., Brun, J.P., Ivers, H., and Cooper, C. (2006). At work but ill: psychosocial work environment and well-being determinants of presenteeism propensity. *Journal of Public Mental Health*, 5, 26-37. Doi 10.1108/17465729200600029

Claes, R. (2011). Employee correlates of sickness presence: A study across four European countries. *Work & Stress*, 25, 224-242. doi 10.1080/02678373.2011.605602

Collins, A., and Cartwright, S. (2012). Why come into work ill? Individual and organizational factors underlying presenteeism. *Employee Relations*, 34, 429-442. doi 10.1108/01425451211236850

Conway, P.M., Hogh, A., Rugulies, R., and Hansen, A.M. (2014). Is sickness presenteeism a risk factor for depression? A Danish 2-year follow-up study. *J Occup Environ Med*, 56, 595-603. doi 10.1097/JOM.0000000000000177

Crowe, M., and Sheppard, L. (2011). A review of critical appraisal tools show they lack rigor: alternative tool structure is proposed. *Journal of clinical epidemiology*, 64, 79-89. oi:10.1016/j.jclinepi.2010.02.008

Dellve, L., Hadzibajramovic, E., and Ahlborg, G., Jr. (2011). Work attendance among healthcare workers: prevalence, incentives, and long-term consequences for health and performance. *Journal of Advanced Nursing*, 67, 1918-1929. doi 10.1111/j.1365-2648.2011.05630.x

Demerouti, E., Le Blanc, P.M., Bakker, A.B., Schaufeli, W.B., and Hox, J. (2009). Present but sick: a three-wave study on job demands, presenteeism and burnout. *Career Development International*, 14, 50-68. doi 10.1108/13620430910933574

Fayers, P.M., and Sprangers, M.A. (2002). Understanding self-rated health. *The Lancet*, 359, 187-188. DOI: 10.1016/S0140-6736(02)07466-4

Gustafsson, K., and Marklund, S. (2011). Consequences Of Sickness Presence And Sickness Absence On Health and Work Ability: A Swedish Prospective Cohort Study. *International*

Journal of Occupational Medicine and Environmental Health, 24, 153-165. doi 10.2478/s13382-011-0013-3

Gustafsson, K., and Marklund, S. (2013). Associations between health and combinations of sickness presence and absence. *Occup Med (Lond)*, 64, 49-55. doi 10.1093/occmed/kqt141

Hackett, R. D., Bycio, P., and Guion, R. M. (1989). Absenteeism among hospital nurses: An idiographic- longitudinal analysis. *Academy of Management Journal*, *32*(2), 424-453. doi: 10.2307/256369

Hansen, C.D., and Andersen, J.H. (2008). Going ill to work - What personal circumstances, attitudes and work-related factors are associated with sickness presenteeism? *Social Science and Medicine*, 67, 956-964. doi 10.1016/j.socscimed.2008.05.022

Hansen, C.D., and Andersen, J.H. (2009). Sick at work-a risk factor for long-term sickness absence at a later date? *Journal of Epidemiology and Community Health*, 63, 397-402. doi 10.1136/jech.2008.078238

Higgins, J.P.T and Green S. (Editors) (2011). Cochrane handbook for systematic reviews of interventions version 5.1. 0. *The Cochrane Collaboration.*

Janssens, H., Clays, E., De Clercq, B., De Bacquer, D., and Braeckman, L. (2013). The relation between presenteeism and different types of future sickness absence. *J Occup Health*, 55, 132-141. doi 10.1539/joh.12-0164-OA

Johns, G. (2010). Presenteeism in the workplace: A review and research agenda. *Journal of Organizational Behavior*, 31, 519-542. doi 10.1002/job.630

Khan, K.S., Ter Riet, G., Glanville, J., Sowden, A.J., and Kleijnen, J. (2001). *Undertaking systematic reviews of research on effectiveness: CRD's guidance for carrying out or commissioning reviews*: NHS Centre for Reviews and Dissemination.

Kivimaki, M., Head, J., Ferrie, J.E., Hemingway, H., Shipley, M.J., Vahtera, J., et al.,(2005). Working while ill as a risk factor for serious coronary events: the Whitehall II study. *American Journal of Public Health*, 95, 98-102. doi 10.2105/AJPH.2003.035873

Kristensen, T. S. (1991). Sickness absence and work strain among Danish slaughterhouse workers: an analysis of absence from work regarded as coping behaviour. *Social science and medicine*, *32*(1), 15-27. doi 10.1080/09585190600878527

Leineweber, C., Westerlund, H., Hagberg, J., Svedberg, P., Luokkala, M., and Alexanderson, K. (2011). Sickness presenteeism among Swedish police officers. *J Occup Rehabil*, 21, 17-22. Doi 10.1007/s10926-010-9249-1 doi 10.1007/s10926-010-9249-1

Liberati, A., Altman, D.G., Tetzlaff, J., Mulrow, C., Gøtzsche, P.C., Ioannidis, J.P., et al.,(2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *Annals of internal medicine*, 151, W-65-W-94. doi:10.7326/0003-4819-151-4-200908180-00136

Lu, L., Lin, H.Y., and Cooper, C.L. (2013). Unhealthy and present: motives and consequences of the act of presenteeism among Taiwanese employees. *Journal of Occupational Health Psychology*, 18, 406. DOI: 10.1037/a0034331

Lu, L., Peng, S.-Q., Lin, H.Y., and Cooper, C.L. (2014). Presenteeism and health over time among Chinese employees: The moderating role of self-efficacy. *Work & Stress*, 28, 165-178. doi:10.1080/02678373.2014.909904

Roelen, C.A., Jensen, C., Stapelfeldt, C.M., Groothoff, J.W., Nielsen, C.V., and Bultmann, U. (2014). The added value of sickness presenteeism to prediction models for sickness absence. *J Occup Environ Med*, 56, e58-59. Doi 10.1097/JOM.0000000000000219

Sanderson, S., Tatt, I.D., and Higgins, J.P. (2007). Tools for assessing quality and susceptibility to bias in observational studies in epidemiology: a systematic review and annotated bibliography. *International journal of epidemiology*, 36, 666-676. doi: 10.1093/ije/dym018

Schultz, A.B., and Edington, D.W. (2007). Employee health and presenteeism: A systematic review. *Journal of Occupational Rehabilitation*, 17, 547-579. doi 10.1007/s10926-007-9096-x

Severens, J.L., Mulder, J., Laheij, R.J., and Verbeek, A.L. (2000). Precision and accuracy in measuring absence from work as a basis for calculating productivity costs in The Netherlands. *Social Science and Medicine*, 51, 243-249. doi10.1016/S0277-9536(99)00452-9

Taloyan, M., Aronsson, G., Leineweber, C., Magnusson Hanson, L., Alexanderson, K., and Westerlund, H. (2012). Sickness presenteeism predicts suboptimal self-rated health and sickness absence: a nationally representative study of the Swedish working population. *PloS one*, 7, e44721. doi 10.1371/journal.pone.0044721

Van Poppel, M.N., de Vet, H.C., Koes, B.W., Smid, T., and Bouter, L.M. (2002). Measuring sick leave: a comparison of self-reported data on sick leave and data from company records. *Occup Med (Lond)*, 52, 485-490. doi: 10.1093/occmed/52.8.485

Voss, M., Stark, S., Alfredsson, L., Vingård, E., and Josephson, M. (2008). Comparisons of self-reported and register data on sickness absence among public employees in Sweden. *Occup Environ Med*, 65, 61-67. doi:10.1136/oem.2006.031427

Figure 1: Stages in the paper selection process

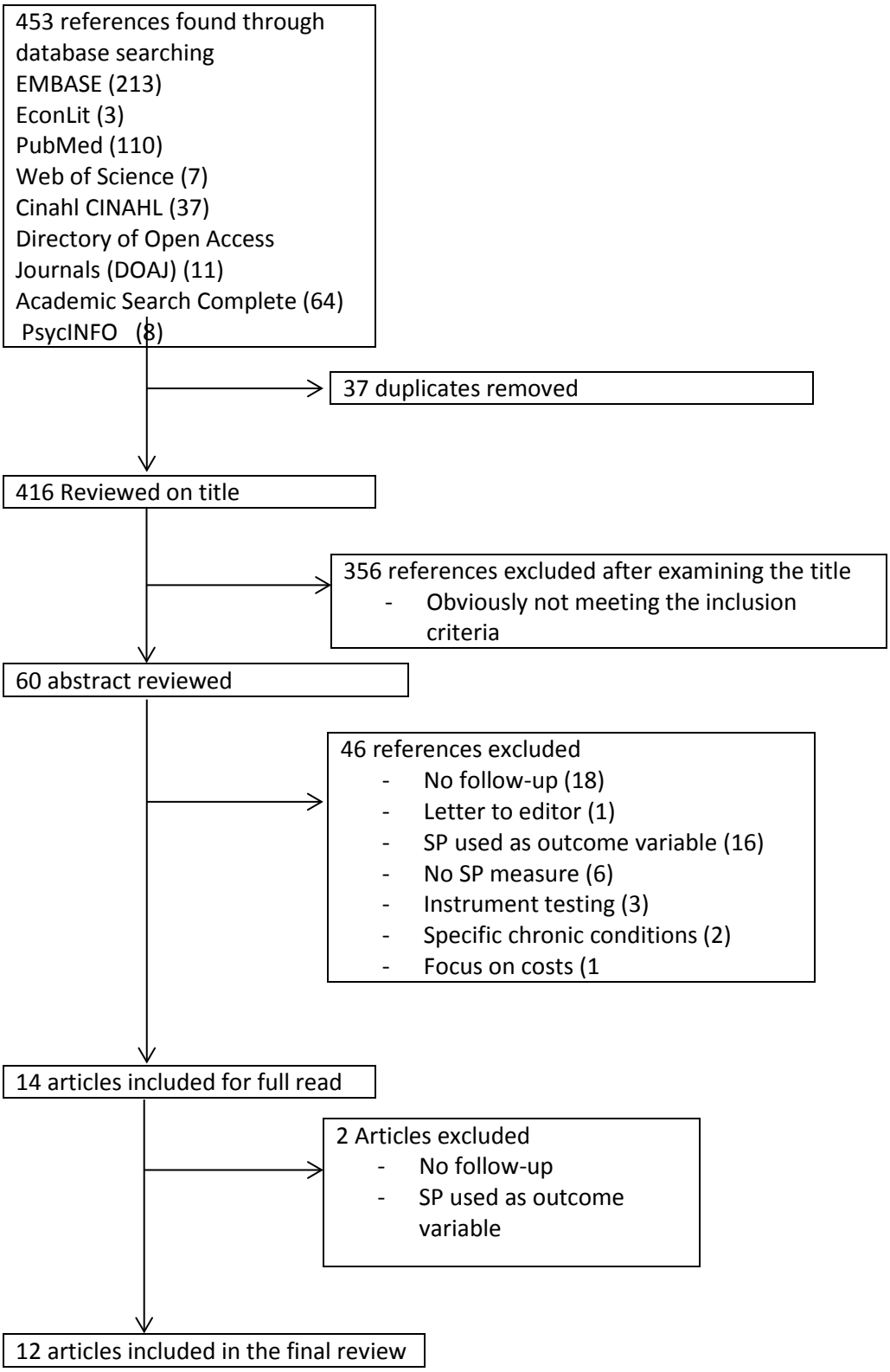

Table 1
Summary of reviewed papers

Author, year, country, final sample size at follow-up (*N*), (response rate)	**Population**	**Statistical analysis and method for inclusion of control variables**	**SP analysis categories**	**Time between follow-up (months)**	**Outcome variable**	**Limitations**	**Strengths**
Bergström et al. (2009a) Sweden. *N*=6,279 (47.7%)	Public and private sector employees	Poisson, BW/FW	0-1 days 2-5 days >5 days	18+36	SRH	Potential risk of recall bias and selection bias (e.g., respondents with continuous sick-leave period longer than 3 months excluded) and attrition bias	2 follow-ups included; validated instruments included; large sample size; non-response analysis carried out
Bergström et al. (2009b) Sweden. *N*=4,687 (45%)	Public and private sector employees	Poisson, BW/FW	0-1 days 2-5 days >5 days	18+36	SA	Potential risk of recall bias and selection bias (e.g., respondents with continuous sick-leave period longer than 3 months excluded) and attrition bias	2 follow-ups included; validated instruments included; large sample size; non-response analysis carried out; recorded sickness absence data
Demerouti et al. (2009) Netherlands. *N*=258 (24.4%)	Nurses in the public health care sector	SE, FX	No/yes	12+18	Burnout	Potential risk of recall bias, selection bias (e.g. small and selective sample) and omitted variable bias	2 follow-ups included; non-response analysis; validated instruments included; SRH included at baseline as a control variable
Hansen & Andersen (2009) Denmark. *N*=11,838 (57.8%)	Random sample of working population	Cox model, TD	0-1 days 2-5 days 6+ days	18	SA	Potential risk of recall bias, selection bias (e.g. exclusion of sick-listed for 10 weeks prior to, and at time of, baseline) and attrition bias	Large and random sample size; moderate response rates; SA(>2 weeks) and SRH at baseline included as control variables
Dellve et al. (2011) Sweden *N*=1,820 (34.3%)	Random sample of public health care sector employees	Cox model, FX	2+ SP, 0-1 SP and 0-7 SA days, the number of SA days taken over previous 2 years	24	SRH and burnout	Potential risk of recall bias, selection bias as low response rate at follow-up and excluded respondents who presented with negative outcomes at baseline (e.g. poor SRH), attrition bias and omitted variable bias	Moderate sample size; includes statistical check on internal consistency using Cronbach's alpha; included validated instruments
Gustafsson & Marklund (2011) Sweden. *N*=2,181 (43.5%)	Random sample of working population	Logreg, BW	None Once 2-5 times >5 times	12+24	SRH, physical complaints and mental wellbeing	Potential risk of recall bias, selection bias and attrition bias	Random and representative sample drawn from registers of Swedish population; 2 follow-up included; moderate response rate and sample size; SA and SRH at baseline included as control variables; validated instruments included

Author, year, country, sample size (*N*), (response rate)	Population	Statistical analysis and method for inclusion of control variables	SP analysis categories	Time between follow-up (months)	Outcome variable	Limitations	Strengths
Taloyan et al (2012) Sweden. *N*=7,445 (not stated)	Representative sample of working population	Logreg, Block	None 1-7 days >7 days	12+24	SRH and SA	Potential risk of recall bias, selection bias (no information on response rates and few participants aged <30 years) and attrition bias	Included 2 follow-up periods; representative sample of the Swedish working population; large sample size; included several control variables including SRH and SA at baseline; models tested using the Hosmer-Lemeshow goodness-of-fit-test
Gustafsson & Marklund (2013) Sweden. *N*=1,886 (37.7%)	Random sample of working population	Mlog, Block	0-1 SP days)/0-6 SA days 7+ SP days 2+ days/0-6 SA days7+ SP days/7+SA days	12+24	SRH, physical complaints and mental wellbeing	Potential risk of recall bias, selection bias, attrition bias and omitted variable bias	Random and representative sample; 2 follow-ups included; moderate response rates and sample size; included several control variables; validated instruments included
Janssen et al. (2013) Belgium. Follow-up *N*=2,876 (30.4%)	Public and private sector employees from seven organisations	Bi-log, TD/BW	0-1 days 2-5 days >5 days	12	SA	Potential risk of recall bias, selection bias and attrition bias	Moderate sample size; Low risk of omitted variables bias; controlled for SRH and SA in all models; non-response analysis carried out
Lu et al. (2013) Taiwan. Follow-up *N*=245 (76.8%)	Taiwanese employees attending university evening classes	Hireg, FX	1=never, 4=>5 times	2	Mental health, physical health and burnout	Potential risk of selection bias (due to small sample size) and omitted variable bias	High response rate; non-response analysis carried out; 2 SP measures included; 6 months recall for SP; confirmatory factor analysis used to test validity of each construct.
Conway et al. (2014) Denmark. *N*=1,271 (51.3%)	Public and private sector employees	Bi-log, TD	None 1-7 days 8+ days	24	Depression	Potential risk of recall and selection bias	Moderate sample size and response rate; non-response analysis carried out; validated instruments included; wide range of control variables
Lu et al (2014) China and Taiwan *N*=345 (90%)	Chinese and Taiwanese employees attending executive training programmes	Hreg, FX	1=never, 4=>5 times	3	Mental health, physical health and burnout	Potential risk of selection bias (due to small sample size) and omitted variable bias	High response rate; non-response analysis carried out; 2 SP measures included; 6 months recall for SP; inclusion of confirmatory factor analysis

Key: **SRH**=Self-rated health, **SA** =sickness absence, **OR**=Odds Ratio, **TD**=theory driven inclusion of control variables, **BW**=Backward selection, **FW**=Forward selection, **FX**=Fixed number of controls, **Block**=Block wise inclusion of controls, **Bi-log**=Bivariate logistic regression, **Hreg**=Hierarchical regression, **Poisson**=Modified Poisson regression, **SE**=Structural equation model, **Logreg**= Multiple logistic regression. **Attrition bias** - no analysis of non-responders. **Omitted variable bias** - variables commonly included as potential confounders of the outcomes not included in the analysis (in this case potential confounders should include SA and SRH as a minimum). **Selection bias** refers to systematic differences between baseline characteristics of participants. Recall bias refers to a systematic error caused by the accuracy that is recalled by respondents regarding past experiences.

Table 2
Summary of the findings

Author, year, country, sample size (*N*), (response rate)	**Population**	**Statistical analysis and method for inclusion of control variables**	**SP analysis categories**	**Time between follow-up (months)**	**Outcome variable**	**Significant findings**
Bergström et al. (2009a) Sweden. Baseline *N*=9,583 (72.8%); final follow-up *N*=6,279 (47.7%)	Public sector: predominantly female in social care and educational sectors; Private sector predominantly male in blue collar industrial sector	Modified Poisson regression. Control: backward and forward selection of control variables	0-1 days 2-5 days >5 days	18+36	SRH measured using a single item from the Short Form-36	For public sector employees at 3 years follow-up: 2-5 days of SP and >5 days were associated with increased risk of poorer SRH (RR 1.38 and 1.90, respectively) for those reporting excellent/good health at baseline. For private sector employees at 3 years follow-up: 2-5 days of SP and >5 days was associated with increased risk of poorer SRH (RR 1.49 and 1.71, respectively) for those reporting fair/poor health at baseline. The same association was found for those reporting good/excellent health at baseline (RR 1.31 and RR 1.79 respectively)
Bergström et al. (2009b) Sweden. Baseline *N*=6,242 (60.1%); final follow-up *N*=4,687 (45%)	Public sector: predominantly female in social care and educational sectors; Private sector predominately male in blue collar industrial workers	Modified Poisson regression Control: Backward and forward selection of control variables	0-1 days 2-5 days >5 days	18+36	SA data from participating organizations' records	For employees in the public sector 2-5 days of SP and >5 days of SP at baseline were associated with increased risk of sick leave after 18 months (RR 1.18 and 1.40, respectively). For employees in the private sector >5 days of SP were associated with increased risk of sick leave after 18 months (RR 1.51) and >5 days at baseline were associated with future sick leave after 3 years (RR 1.49)
Demerouti et al. (2009) Netherlands. Baseline *N*=781 (73.6%); final follow-up *N*=258 respondents (24.4%)	Public sector: Staff nurses across general hospitals	Structural equation modelling. Control: Fixed number of control variables	No/yes	12+18	Dutch version of Maslach Burnout Inventory; 8-item emotional exhaustion subscale; 5-item depersonalisation subscale	SP at T1 (but not at baseline) was associated with depersonalisation (coef. 0.23) at T2
Hansen & Andersen (2009) Denmark. Baseline *N*=14,241 (69.5%); final follow-up *N*=11,838 (57.8%)	Random sample of Danish workforce	Cox proportional hazard regression. Control: Theory driven inclusion of control variables	0-1 days 2-5 days 6+ days	18	SA data from the Danish Ministry of Employment records – only includes spells lasting 2 weeks or more	2-5 and 6+ days of SP were associated with a hazard ratio of 1.25 and 1.53 (respectively) for future SA of at least 2 weeks. 6+ days were also associated with a hazard ratio of 1.74 for future SA of at least 2 months
Dellve et al. (2011) Sweden Baseline *N*=2,624(49.5%); T2 *N*=1,820 (34.3%)	Random sample of Swedish public health care: hospital, primary care, dentistry and administrative staff	Cox proportional hazard regression. Control: 3 controls included	2+ SP, 0-1 SP and 0-7 SA days, the number of SA days taken over previous 2 years	24	SRH measured using a single item question. Burnout measured using the 22-item Shirom-Melamed Burnout Questionnaire	Sickness attendance (2+ SP days) was associated with poor SRH (RR 3.15) and burnout (1.72). Balanced attendance (0-1 SP days and 0-7 SA days) was associated with a decreased risk SRH (RR 0.26) and burnout (RR 0.62)

Author, year, country, sample size (*N*), (response rate)	Population	Statistical analysis and method for inclusion of control variables	SP analysis categories	Time between follow-up (months)	Outcome variable	Significant findings
Gustafsson & Marklund (2011) Sweden. Baseline: *N*=2,493 (49.7%); final follow-up *N*=2,181 (43.5%)	Random sample of 5,009 employees, aged 25- 50 years, from Swedish Working Life Cohort	Multiple logistic regressions. Control: block-wise inclusion of control variables	Never Once 2-5 times >5 times	12+24	SRH measured using a single item question; 5-item physical complaints scale; 10-item mental wellbeing scale; 5-item Work Ability Index (WAI) scale	1 episode of SP was associated with physical complaints (OR 1.60)* after 12 months 2-5 episodes of SP were associated with poor health (OR 1.67), physical complaints (OR 1.68), low work ability (OR 2.31) and future SA (1.75) after 12 months. >5 episodes of SP were associated with poor health (OR 2.53), physical complaints (OR 2.67), low mental wellbeing (OR 2.16)*, low work ability (OR 3.28) and future SA (OR 1.76)* after 12 months
Taloyan et al (2012) Sweden. Baseline: not stated; Final follow-up *N*=7,445 (not stated)	Swedish working population from Swedish Longitudinal Occupational Study of Health	Multiple logistic regressions. Control: block-wise inclusion of control variables	None 1-7 days >7 days	12+24	SRH measured using a single item question; Self-report SA measured using single item	No significant results in the final model (after adjusting for emotional exhaustion at baseline)
Gustafsson & Marklund (2013) Sweden. Baseline: *N*=2,349 (46.9%); final follow-up *N*=1,886 (37.7%)	Random sample of 5,009 employees, aged 25-50 years, from Swedish Working Life Cohort	Multiple logistic regressions. Control: block-wise inclusion of control variables	Low SP (0-1 days)/low SA (0-6 days) Low SP/high SA (7+ days) High SP (2+ days)/Low SA High SP/High SA	12+24	SRH measured using a single item question; 5-item physical complaints scale; 10 item mental wellbeing scale 5-item Work Ability Index (WAI)	High SP (2+days/high SA (7+ days) was associated with future poor SRH (OR 1.87), low work ability (OR 2.64), physical complaints (OR1.64), poor mental well-being (OR 1.49). High SP (2+ days)/low SA (0-6 days) was associated with poor SRH (OR 1.54), low work ability (OR 1.94) and poor mental wellbeing (OR 1.72) Low SP (0-1 days of SP)/high SA (7+ days) was associated with low work ability (OR 1.64)
Janssen et al. (2013) Belgium. Baseline: *N*=2,983 (30.4%); final follow-up *N*=2,876	Employees, aged 30-55 years from seven Belgian companies: including public administration, health care, social work, and manufacturing	Bivariate logistic regression. Control: theory driven and backward selection of control variables	0-1 days 2-5 days >5 days	12	SA data from participating organizations' records	High rates (>5 days) of SP were associated with long spells (>15 consecutive days) of SA (men OR 2.73, women OR 2.40 , short (1-3 days) SA spells (men OR 2.38 women OR 1.90) and associated with high absence frequency (women OR 2.38) Moderate rates (2-5 days) of SP were associated with long SA spells (15 consecutive days) (men OR 1.90)
Lu et al. (2013) Taiwan. Baseline: N=309 (96%); follow-up N=245 (76.8%)	Full-time Chinese employees in Taiwan, from different organizations, attending university evening classes	Hierarchical regression with difference scores. Control: Fixed number of control variables	1=never, 4=>5 times	2	Occupational Stress Indicator: 12-item mental health sub-scale and 6-item physical health scale; Chinese version of Maslach Burnout Inventory: 9-item exhaustion scale	SP was negatively associated with respondents' mental health (coef. -.29) physical health (coef. -.32), exhaustion (coef. .27) at two months follow-up

Author, year, country, sample size (*N*), (response rate)	Population	Statistical analysis and method for inclusion of control variables	SP analysis categories	Time between follow-up (months)	Outcome variable	Significant findings
Conway et al. (2014) Denmark. Baseline: *N*=3,363 (45.7%); follow-up *N*=1,271 (51.3%)	22 private and 38 public sector Danish organisations with 25 plus employees	Bivariate logistic regression. Control: Theory driven inclusion of control variables.	None 1-7 days 8+ days	24	10-item Major Depression Inventory (MDI)	8+ SP episodes at baseline were associated with an increased risk of future depression (OR 2.45) amongst non-depressed respondents (at baseline)
Lu et al (2014) China and Taiwan Baseline: *N*=383 respondents, follow-up *N*=345 (90%)	Working adults in mainland China and Taiwan, from different organizations, attending executive training programmes in two universities in Taiwan and Beijing	Hierarchical regression analysis. Control: fixed group of control variables	1=never, 4=>5 times	3	Occupational Stress Indicator: 12-item mental health sub- scale and 6-item physical health scale; Chinese version of Maslach Burnout Inventory: 9-item exhaustion scale	No significant results in the final model

*significant at the 5% level (Authors used 1% level)Key: Self-rated health (SRH).