

Measuring sick leave: a comparison of self-reported data on sick leave and data from company records

M. N. M. van Poppel^{*}, H. C. W. de Vet^{*}, B. W. Koes[†], T. Smid^{*‡} and L. M. Bouter^{*}

^{*}*Institute for Research in Extramural Medicine, VU University Medical Center, Amsterdam;*

[†]*Department of General Practice, Erasmus University, Rotterdam; and* [‡]*Health Safety and Environment, KLM Royal Dutch Airlines, Schiphol Airport, The Netherlands*

The objective of this study was to compare sick leave data obtained from questionnaires with data from company records. During a period of 12 months, questionnaires were completed monthly for 6 months and then at 9 and 12 months. The sensitivity and specificity of questionnaires for detecting an episode of sick leave were determined, using the company records as a reference standard. In addition, the duration of sick leave episodes reported in the two data sets was compared. In this analysis, company records were not assumed to be superior, and agreement was assessed with intraclass correlation coefficients (ICCs). The sensitivity of questionnaires for detecting an episode of sick leave was 55% (95% CI = 0.50–0.60) and the specificity 83% (95% CI = 0.72–0.94). The ICC for all episodes was 0.58 (95% CI = 0.47–0.67). The only satisfactory ICC (0.87; 95% CI = 0.74–0.93) was found for the questionnaires at 9 and 12 months. No large systematic differences were found between the duration of episodes reported in the two data sets. In conclusion, in our study, the sensitivity of questionnaires for detecting an episode of sick leave was very low. Furthermore, when episodes were recalled, there was little agreement on the duration of the episode between questionnaire data and data in the company records. Based on these results and considering the risk of missing questionnaires, data on sick leave gathered from company records are clearly preferable as an outcome measure in research.

Key words: Agreement; company records; intraclass correlation coefficients; questionnaires; sensitivity; sick leave; specificity.

Received 20 February 2002; revised 6 August 2002; accepted 21 August 2002

Introduction

In epidemiological studies in an occupational setting and in cost-effectiveness studies, return to work, sick leave or work disability are often important outcome measures. These data are usually collected by self-administered questionnaires or by interviews. Given the importance of sick leave as an outcome measure in research, it is surprising that little attention has been paid to the

accuracy of self-reported sick leave. In studies conducted at the workplace, the opportunity often exists to assess the agreement between self-reported sick leave and sick leave registered in company records.

To our knowledge, only a few studies have taken this opportunity [1–5]. Fredriksson *et al.* [1], Agius *et al.* [2] and Burdorf *et al.* [3] all reported a high specificity of a single questionnaire for detecting workers taking sick leave due to back pain during a certain time period (96, 92 and 97%, respectively), using company records as the reference standard. However, the sensitivities that were reported in these studies were lower and varied (76, 68 and 88%, respectively). This variance in reported

Correspondence to: Ms M. N. M. van Poppel, PhD, Department of Social Medicine, Institute for Research in Extramural Medicine, VU University Medical Center, Van der Boechorststraat 7, 1081 BT Amsterdam, The Netherlands. Tel: +31 20 4449697; fax: +31 20 4448387; e-mail: MNM.van_Poppel.EMGO@Med.VU.nl

sensitivity could be due partly to the length of the period that the questionnaire addressed, which was the previous 4 years [1], 12 months [2] and 6 months [3], respectively. Van der Weide [4], who used repeated monthly questionnaires, found that the number of reported days of sick leave due to back pain differed substantially between company records and the self-reported data, especially short episodes of sick leave, which were underestimated in the self-reported data. Severens *et al.* [5], however, did not report systematic differences between registered and reported sick leave data.

In the present study, monthly questionnaires were used to gather data on sick leave during a 6 month study period. In the subsequent 6 months, two questionnaires, each covering 3 months, were distributed. In this paper, analyses are presented in which data on sick leave gathered with the questionnaires are compared with sick leave data from the company records.

Methods

Originally, our study was designed as a randomized controlled trial with three intervention groups (lumbar support and education, education only, and lumbar support only) and a control group. The interventions, however, had no effect on the incidence of back pain or sick leave due to back pain. The results of the intervention study have been reported previously [6].

Subjects

Workers from the Cargo Department of a major Dutch airline company at Schiphol Airport were recruited. All workers involved with manual material handling were invited to participate. Typical tasks of the workers included the loading and unloading of cargo pallets, as well as transportation and sorting of cargo, both manually and with a forklift truck. Workers who had a permanent (partial) work disability were excluded from the study. Of 380 eligible subjects, 20 (5%) refused to participate and 312 (82%) completed the baseline measurements. Because of holidays and a high workload, workers were not always available for the study. Consequently, baseline measurements were missing for 48 (13%) workers. Workers personally received information about the procedures of the study and were enrolled after giving their consent.

Outcome assessment

Questionnaires

During the 6 month study period, the workers received a monthly questionnaire on the occurrence of sick leave and low back pain. Subjects were asked if they had taken sick leave in the past month (yes/no) and, if 'yes', the number of days with sick leave. They were asked the same specifically for sick leave because of back pain (see Appendix). All subjects worked full-time, therefore no correction for part-time work was necessary. In the additional 6 month follow-up period, the same questionnaire on sick leave was filled in twice (at 9 and 12 months after randomization, respectively), each questionnaire covering a 3 month period (Figure 1).

Company medical records

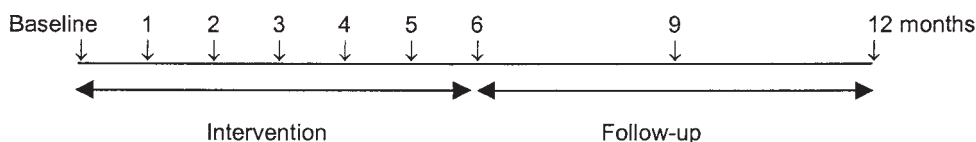
Data on sick leave were also collected from the company records for the total 12 months of the follow-up period. Not all workers were registered in the company records: about one-sixth of the workers participating in the study were temporary workers, who had no records at the company. No sick leave data could be obtained from the temporary employment agency, so the temporary workers were therefore not included in the analyses. On the first day lost from work, the worker calls in to report sickness absence and states the reason for it. This is recorded by a telephone assistant. After 2 or 3 days, a company nurse contacts the worker, and his/her diagnosis is recorded in the records as well. After ~2 weeks of sickness absence, workers are required to visit the company physician, who also registers a diagnosis. The latest diagnosis available for an episode of sick leave was used in this study. Return to work should be reported by the workers on their first day back at work, or the first day on which they are recovered if this is during a weekend or holiday.

Statistical analysis

In the analyses, only episodes in periods for which both company record data and questionnaires were available were included.

First, the accuracy of the number of episodes of sick leave stated in the questionnaires was studied. For this analysis, the medical registration was considered to be the reference standard. The sensitivity and specificity for

Figure 1. Time schedule of measurements. The arrows indicate when a questionnaire was administered.



detecting an episode of sick leave with questionnaires was calculated.

In the comparison of the duration of episodes in the questionnaire data and the company records, only the 176 episodes recalled in the questionnaires were included. Furthermore, for the duration of sick leave episodes, no superiority of the data from the company records was assumed, because the registered duration would not always be correct. Workers sometimes forget to report their return to work on their first day back at work, or their day of recovery during a holiday or weekend. The duration of sick leave episodes in the company records will therefore be overestimated in some cases. Since neither data set was assumed to be superior, the agreement between the duration of sick leave episodes in the two data sets was determined. For assessing the agreement between the duration of sick leave episodes according to questionnaire data and that according to the registration, intraclass correlation coefficients (ICCs) were calculated [7]. The ICC was defined as the ratio of variance among subjects over the total variance, which ranges from 0 (= no reliability) to 1 (= perfect reliability). The ICC was calculated using the two-way random effects model [8].

Because of potentially different recall of episodes of sick leave, analyses were performed separately for the monthly and 3-monthly questionnaires, and for short (≤ 7 days) and long (> 7 days) episodes of sick leave. Because the data were gathered in a randomized study focused on the prevention of low back pain, the recall for sick leave due to back pain could be different from that for sick leave due to other reasons. Therefore, analyses were also repeated regarding episodes due to back pain only.

Results

The 312 workers participating in the intervention study were almost all males (99%), with a mean age of 35.1 years ($SD = 7.8$). The workers had been employed at the Cargo Department for 6.3 years ($SD = 5.9$) on average. At least one questionnaire was filled in by 287 (92%) workers during the follow-up period, of whom 46 (16%) were not in the company records because they were temporary workers (Figure 2). The remaining 241 workers in the company records were responsible for 646 sick leave episodes in 12 months, according to the medical records. Of these 646 episodes, 176 (27%) were also reported in the questionnaires, 144 (22%) were not recalled by workers, and for 326 (50%) episodes no questionnaire was filled in during that period of the follow-up. Of the 241 workers, 35 (15%) did not take sick leave during the follow-up period according to the company records (Figure 2).

In the analyses presented in this paper, the 326 episodes for which no questionnaire was filled in during

that period of the follow-up were excluded. Based on the remaining 320 episodes, and assuming that the medical registration is 100% accurate in registering whether a worker takes sick leave or not, the sensitivity of the questionnaires was 55% (176/320; 95% CI = 0.50–0.60). Seven episodes of sick leave reported in the questionnaires were not registered in the medical records. The specificity was 83% (34/41; 95% CI = 0.72–0.94).

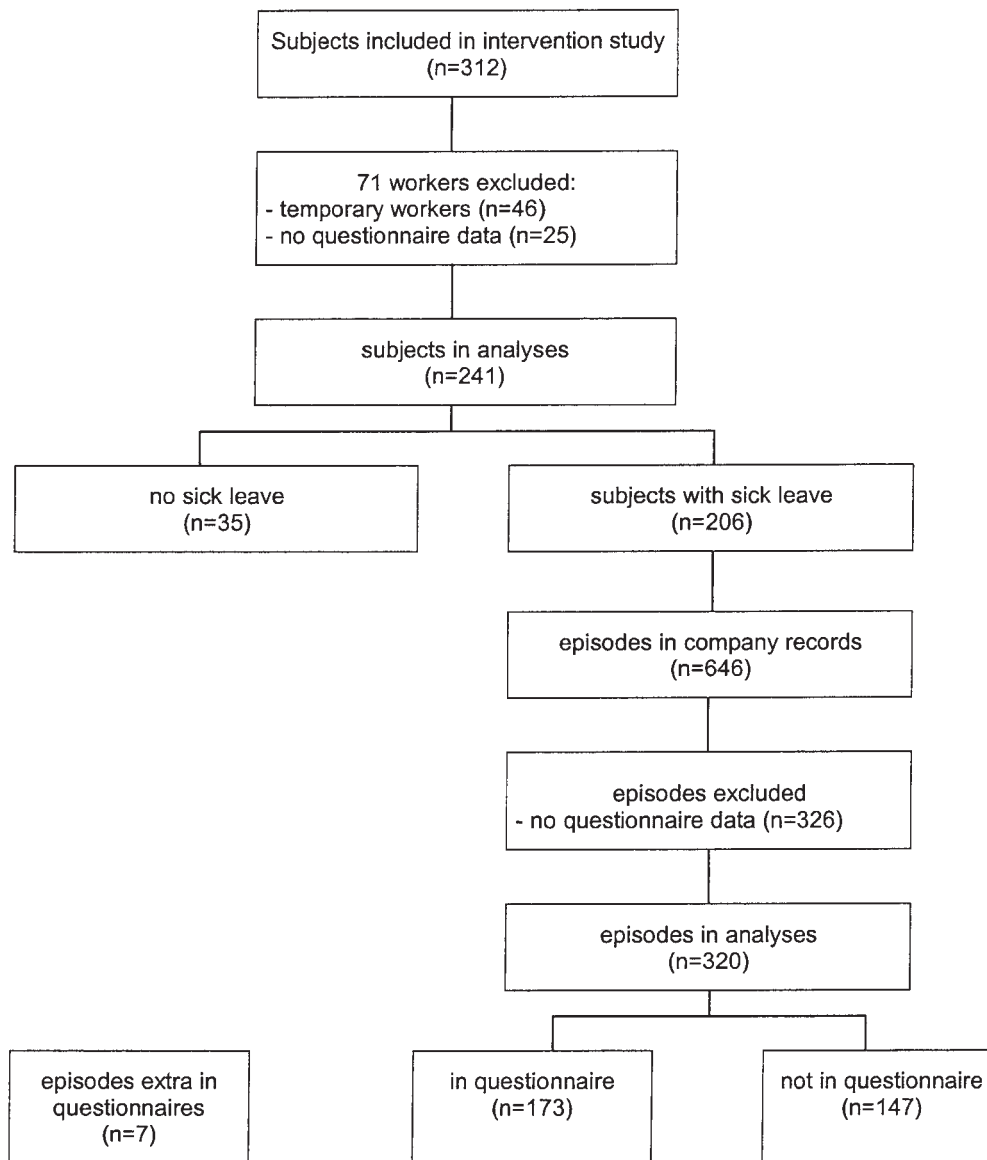
The agreement of the duration of episodes of sick leave reported in the questionnaires with the duration in the company records was assessed. These analyses were based on 173 episodes that were recalled by workers and for which a duration was reported (Table 1). The ICC for duration for all episodes of sick leave was 0.58 (95% CI = 0.47–0.67). The ICC for duration for the 3-monthly questionnaires was higher than that for the monthly questionnaires (0.87 versus 0.51). The ICC for duration for episodes of > 7 days was lower than that for episodes of ≤ 7 days (0.33 versus 0.57). The duration of episodes of sick leave due to low back pain was longer than the average duration of sick leave episodes for all diseases and the ICC of 0.34 was similar to the ICC of 0.33 for episodes of > 7 days (Table 1).

Discussion

One of the limitations of this study is the number of missing questionnaires: questionnaire data were missing for 50% of all sick leave episodes registered in the company records. Because we did not want this to influence the comparison between the sick leave data from the company records and those from the questionnaires, we excluded all episodes for which no questionnaire data were available. The mean duration of these episodes was identical to the mean duration of those episodes included in the analyses. There thus appears to have been no selection bias based on the duration of the sick leave episode. However, missing questionnaires is a phenomenon that should be taken into account when using sick leave from questionnaire data as an outcome measure in a clinical trial.

In the present study, the company records were used as the reference standard for calculating the sensitivity and specificity of the questionnaires, but it is not certain that all episodes of sick leave are indeed covered by the company records. In addition, the diagnoses in the records may not always be completely valid, especially when the reason for sick leave is stated by the worker. It may thus be possible that in the analyses on sick leave due to back pain, some of the sickness absence episodes have been misclassified in the company records.

The rationale for gathering data during the follow-up period by repeated questionnaires was to keep recall bias minimal. However, minimizing recall bias does not seem to have been very effective in our study. In the returned

Figure 2. Flowchart of subjects and episodes of sick leave.**Table 1.** The mean duration of sick leave episodes according to the company records and questionnaires, the mean difference in duration between the company records and questionnaire data, and the ICCs for the duration

	<i>Mean duration in days (median)</i>		<i>Mean difference (95% CI)</i>	<i>ICC (95% CI)</i>
	<i>Company records</i>	<i>Questionnaire</i>		
All episodes (<i>n</i> = 173)	10.1 (6)	8.1 (5)	2.0 (−16.9–20.9)	0.58 (0.47–0.67)
Monthly questionnaire (<i>n</i> = 142)	10.3 (6)	7.5 (5)	5.4 (−24.4–35.3)	0.51 (0.37–0.62)
Three-monthly questionnaire (<i>n</i> = 31)	9.4 (5)	10.9 (4)	0.1 (−3.8–4.0)	0.87 (0.74–0.93)
Episodes due to back pain (<i>n</i> = 26) ^a	17.8 (9)	13.5 (9)	4.3 (−31.2–39.8)	0.34 (−0.03–0.64)
Episodes of ≤7 days (<i>n</i> = 110)	4.2 (4)	4.1 (4)	2.8 (−17.0–22.6)	0.57 (0.43–0.68)
Episodes of >7days (<i>n</i> = 63)	20.7 (15)	15.2 (14)	−1.5 (−13.4–10.4)	0.33 (0.10–0.53)

^aThese episodes are due to back pain according to the company records. Episodes due to back pain according to the questionnaires only (*n* = 10) were not included.

questionnaires, workers recalled only 55% of the sick leave episodes. This low recall percentage in our study may partly be caused by the fact that workers were 'rewarded' for answering negatively on the question about sick leave, because they could skip several questions if they had had no sick leave in the previous period. The population under study was not used to receiving and completing written information. Therefore, another possibility is that these workers were not very motivated for filling in questionnaires, which is illustrated by the number of questionnaires missing. In addition, it is possible that workers did not really try to remember sick leave episodes when they *did* fill in a questionnaire.

This finding of substantial recall bias using monthly questionnaires is in contrast to the results of Severens *et al.* [5]. In their study among office workers, a questionnaire was used for measuring the number of days of sick leave during five different recall periods. The questionnaire data were compared with data on sick leave from the company records. Severens *et al.* [5] reported that recall periods of 6 and 12 months might lead to recall bias, whereas recall periods of ≤ 2 months might avoid such bias. The difference in findings between their study and the one presented here may be explained by the different study populations (blue-collar workers versus office workers) and the incidence of sick leave, which was high in this study and low in the study of Severens.

All three previous studies reported higher sensitivities for detecting sick leave due to back pain. Fredriksson *et al.* [1] reported a sensitivity of 76% for a single questionnaire covering a period of 4 years, Agius *et al.* [2] reported a sensitivity of 68% for one covering 12 months and Burdorf *et al.* [3] reported a sensitivity of 79% for a questionnaire covering a period of 6 months. However, these studies did not assess the sensitivity for detecting single episodes, but rather assessed the ability of questionnaires to detect whether a subject had had sick leave during the follow-up period or not, regardless of the number of episodes. If a subject forgot one of a number of episodes of sick leave, this would have had no consequence for the reported sensitivity. This could explain why the levels of sensitivity that the above authors reported were higher than in our study. Similar analyses were performed with our data, using only data from workers who had filled in all questionnaires during the intervention period. In this analysis, the sensitivity was 71% and the specificity 100% for detecting whether or not workers had taken sick leave during the 6 month period (data not shown). This result is comparable to the results of the previous studies.

The ICC for all episodes of 0.58 indicates that the agreement between the duration reported in the company medical records and the questionnaires is disappointingly

low. The lowest ICCs were found for longer episodes of sick leave (episodes of >7 days and episodes due to back pain). The only satisfactory ICC (0.87) was found for the agreement between the duration in the company records and the 3-monthly questionnaires.

This finding is in line with the results of the study of Severens *et al.* [5]. In that study, an effort was made to find the 'optimal' recall period for measuring sick leave. The authors suggested that the optimally efficient recall period is somewhere between 2 and 6 months, and, based on their results, recommended a recall period of 2 months.

In conclusion, in our study, the sensitivity of questionnaires for detecting an episode of sick leave was very low. Furthermore, when episodes were recalled, there was little agreement on the duration of the episode between questionnaire data and the data in the company records. Based on these results and considering the risk of missing questionnaires, data on sick leave gathered from company records are clearly preferable as an outcome measure in a trial. This is, of course, dependent on the company having a reliable registration system of sick leave episodes.

References

1. Fredriksson K, Toomingas A, Torgen M, Thorbjornsson CB, Kilbom A. Validity and reliability of self-reported retrospectively collected data on sick leave related to musculoskeletal diseases. *Scand J Work Environ Health* 1998; **24**: 425–431.
2. Agius RM, Lloyd MH, Campbell S, Hutchison P, Seaton A, Soutar CA. Questionnaire for the identification of back pain for epidemiological purposes. *Occup Environ Med* 1994; **51**: 756–760.
3. Burdorf A, Post W, Bruggeling T. Reliability of a questionnaire on sickness absence with specific attention to absence due to back pain and respiratory complaints. *Occup Environ Med* 1996; **53**: 58–62.
4. Van der Weide WE. Quality of occupational rehabilitation for low-back pain: measurement of process and outcome. Thesis, Coronel Institute, University of Amsterdam, 1998.
5. Severens JL, Mulder J, Laheij RJF, Verbeek ALM. Precision and accuracy in measuring absence from work as a basis for calculating productivity costs in The Netherlands. *Social Sci Med* 2000; **51**: 243–249.
6. Van Poppel MNM, Koes BW, van der Ploeg GE, Smid T, Bouter LM. Lumbar supports and education for the prevention of low back pain in industry: a randomized controlled trial. *J Am Med Assoc* 1998; **279**: 1789–1794.
7. Rankin G, Stokes M. Reliability of assessment tools in rehabilitation: an illustration of appropriate statistical analyses. *Clin Rehabil* 1998; **12**: 187–199.
8. Fleiss JF. *The Design and Analysis of Clinical Experiments*. London: John Wiley & Sons, 1986.

Appendix. Part of the questionnaire concerning back pain and sick leave

1. Did you experience back pain in the past four (twelve) weeks? Yes → continue with questions on duration, location, severity of back pain, etc. (not shown)
 No → continue with question 2

2. Did you take sick leave in the past four (twelve) weeks? Yes → continue with question 3
 No → continue with questions on work tasks (not shown)

3. How many days did you take sick leave in the past four (twelve) weeks? ... days

4. Did you take sick leave due to back pain in the past four (twelve) weeks? Yes
 No

If yes:

How many days did you take sick leave due to back pain in the past four (twelve) weeks? ... days
