Inter-rater reliability of direct observations of the physical and psychosocial working conditions in eldercare

An evaluation in the DOSES project

Karstad, Kristina; Rugulies, Reiner; Skotte, Jørgen; Munch, Pernille Kold; Greiner, Birgit A; Burdorf, Alex; Søgaard, Karen; Holtermann, Andreas

Published in:
Applied Ergonomics

DOI:
10.1016/j.apergo.2018.01.004

Publication date:
2018

Document version
Publisher's PDF, also known as Version of record

Applied Ergonomics 69 (2018) 93–103

Contents lists available at ScienceDirect

Applied Ergonomics

journal homepage: www.elsevier.com/locate/apergo

Inter-rater reliability of direct observations of the physical and psychosocial working conditions in eldercare: An evaluation in the DOSES project

Kristina Karstad[a,b,*], Reiner Rugulies[a,c,d], Jørgen Skotte[a], Pernille Kold Munch[a], Birgit A. Greiner[e], Alex Burdorf[f], Karen Søgaard[b,g], Andreas Holtermann[a,b]

[a] National Research Centre for the Working Environment (NRCWE), Lersø Parkallé 105, 2100 Copenhagen Ø, Denmark
[b] Department of Sports Science and Clinical Biomechanics, University of Southern Denmark, Campusvej 55, 5230 Odense M, Denmark
[c] Department of Public Health, University of Copenhagen, Øster Farimagsgade 5, 1014 Copenhagen K, Denmark
[d] Department of Psychology, University of Copenhagen, Øster Farimagsgade 2A, 1353 Copenhagen K, Denmark
[e] School of Public Health, University College Cork, Western Road, Cork, Ireland
[f] Department of Public Health, Erasmus University Medical Centre, P.O. Box 2040, 3000 CA, Rotterdam, The Netherlands
[g] Department of Clinical Research, University of Southern Denmark, Winsløwparken 19, 5000 Odense C, Denmark

ARTICLE INFO

Keywords:
Nursing home
Healthcare
Patient handling activities

ABSTRACT

The aim of the study was to develop and evaluate the reliability of the "Danish observational study of eldercare work and musculoskeletal disorders" (DOSES) observation instrument to assess physical and psychosocial risk factors for musculoskeletal disorders (MSD) in eldercare work. During 1.5 years, sixteen raters conducted 117 inter-rater observations from 11 nursing homes. Reliability was evaluated using percent agreement and Gwet's AC1 coefficient. Of the 18 examined items, inter-rater reliability was excellent for 7 items (AC1 > 0.75) fair to good for 7 items (AC1 0.40–0.75) and poor for 2 items (AC1 0–0.40). For 2 items there was no agreement between the raters (AC1 < 0). The reliability did not differ between the first and second half of the data collection period and the inter-rater observations were representative regarding occurrence of events in eldercare work. The instrument is appropriate for assessing physical and psychosocial risk factors for MSD among eldercare workers.

1. Introduction

Musculoskeletal disorders (MSD) are highly prevalent among eldercare workers (Davis and Kotowski, 2015; Luime et al., 2004). Correspondingly rates of sickness absence (Andersen et al., 2012) and premature retirement from the labor market (Jensen et al., 2012) are also high in this job group.

Eldercare workers' primary task is to take care of the residents, which often includes manual handling activities like lifting, repositioning, turning, pulling on/off compression stockings and pushing and pulling residents in different portable chairs. These manual handling tasks can be physically demanding for the eldercare worker, and potentially increase the risk for MSD (Lagerström et al., 1998; Trinkoff et al., 2003), which may result in sickness absence (Andersen et al., 2012) and premature retirement (Jensen et al., 2012).

Caring for residents not only includes satisfying physical needs but also emotional caring activities. Caring may include both verbal and physical interactions between the eldercare worker and the resident that can be of both positive and negative character. A systematic review and meta-analysis of 54 cohort studies showed that adverse psychosocial working conditions were prospectively associated with risk of MSD (Hauke et al., 2011). With regard to care workers, two recent studies reported that violence and aggression of the resident towards the care workers predicted risk of MSD and sickness absence among eldercare workers (Aagestad et al., 2014; Miranda et al., 2014, 2011).

Effective workplace surveillance, risk evaluation and preventive interventions for eldercare work rely on reliable measurements of physical and psychosocial factors in the care of elderly. Self-reported assessment of these factors may be imprecise and biased (Gupta et al., 2016; Jakobsen et al., 2016; Koch et al., 2016; Kwak et al., 2011; Prince et al., 2008). Therefore, observation methods have been developed and applied for assessing these exposures in elderly care (Jakobsen et al., 2016, 2015; Johnsson et al., 2004; Park et al., 2009). However, even though the manual handling activities and psychosocial interaction in the caring situation of the elderly often occur in parallel and may impact upon each other, no previous observation instruments have been developed to assess both factors simultaneously.

The "Danish observational study of eldercare work and

* Corresponding author. National Research Centre for the Working Environment (NRCWE), Lersø Parkallé 105, 2100 Copenhagen Ø, Denmark.
E-mail address: kka@nrcwe.dk (K. Karstad).

https://doi.org/10.1016/j.apergo.2018.01.004
Received 19 April 2017; Received in revised form 31 October 2017; Accepted 6 January 2018
Available online 30 January 2018

musculoskeletal disorders" (DOSES) is a large prospective study in 126 wards in 20 nursing homes with the main aim to investigate the independent and combined contribution of physical and psychosocial working conditions to the occurrence of MSD and its consequences among Danish eldercare workers. We developed an observation instrument for the simultaneous assessment of physical and psychosocial risk factors for MSD by direct observations of eldercare work.

The main purpose of this article is to examine the inter-rater reliability of the DOSES observation instrument. In direct observations of daily work over longer time periods, the agreement between raters is particularly vulnerable to vary with the exposures being observed, the training and experience of the raters and the characteristics of the job (Park et al., 2009; Voskuijl and van Sliedregt, 2002). Furthermore, we investigate whether the inter-rater reliability depends on the exposure type and improves by time throughout the data collection period of 1.5 years.

2. Material and methods

We conducted the reliability evaluation in a sample of the study population in DOSES from September 2013 to December 2014. The DOSES observation instrument is based on direct observations of caring activities involving both residents and eldercare workers.

2.1. Development of the instrument

In 2012, we established a working group consisting of two researchers from the psychosocial work environment field, two researchers from the field of physical activity and demands at work and one occupational therapist. The working group collaborated with three experienced researchers, from June 2012 to April 2013, in the development of the DOSES observation instrument. The DOSES observation instrument was based on two earlier observational instruments; one instrument for observing psychosocial work environment in eldercare (Jakobsen et al., 2016, 2015) and one instrument for observing ergonomic factors (Koppelaar et al., 2012). The two instruments were developed by two experts who were also involved in the development of the DOSES observation instruments.

In a recent study that used one of the earlier observational instruments the authors found that frequent social interactions between care workers and residents were associated with higher depressive symptoms among care workers (Jakobsen et al., 2016). The design of that previous study allowed the authors only to analyze the frequency of the social interactions but not the content of the interaction. These concepts were based on the concept of 'emotional labour' as formulated by Zapf (2002) from an action regulation theoretical perspective. Emotional labour describes the process of managing feelings to fulfill the emotional requirements of a job and of clients or patients. Emotion regulation puts additional demands on workers with positive or negative effects. Because depressive symptoms and MSD are correlated with each other (Clausen et al., 2013; del Campo et al., 2017), we considered it possible that social interactions between care workers and residents, and in particular social interactions with a negative content (e.g. hostile behavior of the resident), may be related to risk of MSD among care workers. Therefore, we took our point of departure in the earlier instrument (Jakobsen et al., 2016, 2015) and further refined the instrument in a way that allowed us not only to measure frequency but also positive and negative content of social interactions between care workers and residents.

In April and May 2013, we conducted a pilot study comprising 5 wards from 3 nursing homes on 34 eldercare workers and 112 residents to test the procedures and feasibility of the methods, and the reliability of the observation instrument. After the pilot study, a discussion was held between the research group and the observers to discuss the feasibility of the methods and any obstacles that appeared. The observation instrument was considered to be feasible to use. It also showed fair to good inter-rater agreement, why only few adjustments were made after the pilot study.

2.1.1. The DOSES observation instrument

The coding and data entry scheme for the DOSES observation instrument was created on a computer using the software Noldus Observer XT 11 (Noldus, Wageningen, The Netherlands). The DOSES data entry scheme was transferred to tablets containing the commercially available software Noldus Observer XT pocket observer. This Noldus Observer software was used for data entry of the real-time inter-rater observations. The overall sampling was continuous in time, giving the opportunity to record both durations (start and stop time) and instantaneous occurrence (point-events) of the registered items.

The definition of "an observation" in this study is the observed continuous sequence of caring activities involving both resident and eldercare worker. The observation started when an eldercare worker entered the room of a resident, and the observation stopped when the eldercare worker finalized the caring activities of the resident and left the room. Within a single observation, the observer reported every event that occurred.

The overview of the DOSES observation instrument is presented in Table 1. Overall, the observation instrument was composed of 26 items for observation. For item 3, 10–17 and 25–26, additional information (referred to as descriptive factors), was added to provide more descriptive information to the specific item.

The 26 items were defined as either a "point event" or "state event" referring to how the events of the items were registered and thus the information they provide. A "point event" was registered at a single time point, providing information of the occurrence of an event. A "state event" was registered over time, containing information of both the occurrence and the duration of the event. The duration of the event was based on either manual registration of a start- and stop-time (referred to as "Start-Stop") or with manual start-time and automatic stop-time when a new item was registered (referred to as "Mutually exclusive and exhaustive").

Part 1 of the observation instrument (item 1 to 9 in Table 1) recorded the setting of the observation including caring activity in day shift, caring activity in evening shift, feeding situation, other situation and denied access to the room. The reason for denied access of the rater to the room of the resident (coded as descriptive factor 1 in the instrument) was given by the eldercare worker. These were registered as "Mutually exclusive and exhaustive". The rater registered manually with "start-stop" when a colleague or another person was present during the observation. Other occupational hazards (i.e. second-hand smoking or if the eldercare worker had to move furniture) were registered as single point events.

Part 2 of the observation instrument (item 10 to 18 in Table 1) recorded manual handling activities. Lifting a resident was defined as lifting and lowering a resident from one surface to the same or another surface. Repositioning a resident was defined as moving a resident up/down/sideways in bed, assisting the resident in rising to sit on the edge of the bed, or moving the resident forward/backwards on a chair without lifting the resident out of the chair. Turning a resident was defined as rolling a resident from the back position to a side position or vice-versa. It was also registered whether the resident helped substantially during the manual handling activities (defined as at least 25% reduction in physical load for the eldercare worker), whether an assistive device was used or whether a colleague (coded as descriptive factor 1 in the instrument) or others helped with the handling activity (coded as descriptive factor 2 in the instrument). Lifting, repositioning and turning the resident as well as pulling a support stocking up or down, or pushing/pulling a resident in a portable chair were registered as single point events the moment it occurred. Squatting was defined as working position with the knees bent to less than a 90° angle or kneeling on the floor (two merged items from Buchholz et al., 1996), and was registered as a "start-stop" event.

Table 1
Overview of the DOSES observation instrument for assessing physical and psychosocial risk factors for musculoskeletal disorders in eldercare work. The instrument consists of 26 items for observation with up to two descriptive factors which serve as subcategories of the item. Type of event defines the sampling strategies applied.

Item	Descriptive factor 1 (DF1)	Descriptive factor 2 (DF2)	Type of event[a]
Setting/surroundings			
1. Care in day shift			State event/Mutually exclusive and exhaustive
2. Care in evening shift			State event/Mutually exclusive and exhaustive
3. Denied access to resident	a. Emotional reaction b. Physical reaction c. Ethical reason d. Other		State event/Mutually exclusive and exhaustive
4. Feeding			State event/Mutually exclusive and exhaustive
5. Other situation			State event/Mutually exclusive and exhaustive
6. Colleague present			State event/Start-Stop
7. Other present			State event/Start-Stop
8. Smoke in the room			Point event
9. Move furniture			Point event
Manual handling activities			
10. Lifting with support from resident	a. Transfer belt b. Sliding sheet c. Other d. No assistive device	a. With help from colleague b. Without help from colleague	Point event
11. Lifting without support from resident	a. Floor hoist b Ceiling hoist c. Sliding sheet d. Other e. No assistive device	a. With help from colleague b. Without help from colleague	Point event
12. Repositioning with support from resident	a. Sliding sheet b. Draw sheet c. Other d. No assistive device	a. With help from colleague b. Without help from colleague	Point event
13. Repositioning without support from resident	a. Sliding sheet b. Draw sheet c. Other d. No assistive device	a. With help from colleague b. Without help from colleague	Point event
14. Turning with support from resident	a. Draw sheet b. Electric turning sheet c. Other d. No assistive device	a. With help from colleague b. Without help from colleague	Point event
15. Turning without support from resident	a. Draw sheet b. Electric turning sheet c. Other d. No assistive device	a. With help from colleague b. Without help from colleague	Point event
16. Support stockings	a. With assistive device b. Without assistive device		Point event
17. Push/pull resident in portable chair	a. Wheel chair b. Hoist c. Bath/toilet chair d. Other		Point event
18. Squatting			State event/Start-Stop
Psychosocial interactions			
19. Physical appreciation			Point event
20. Verbal appreciation			Point event
21. Physical resistance			Point event
22. Verbal resistance			Point event
23. Physical aggression			Point event
24. Verbal aggression			Point event
Barriers			
25. Impediments	a. Missing supplies/equipment b. Broken supplies/equipment c. Missing colleague d. Other		State event/Start-Stop
26. Interruptions	a. From colleague b. From other resident c. Other		State event/Start-Stop

[a] "Point event" is registered at a single time point providing information of the instantaneous occurrence of an event. "State event/Start-Stop" is registered over time using manual start- and stop-time providing information of the instantaneous occurrence and duration of an event. "State event/Mutually exclusive and exhaustive" is registered over time using manual start-time and automatic stop-time by registration of another item providing information of the instantaneous occurrence and duration of an event.

Part 3 of the observation instrument (item 19 to 24 in Table 1) recorded psychosocial interactions between the care worker and the resident, defined as emotional reactions from the resident that were assumed to be either a resource or a demand for the eldercare worker. Resources included "physical appreciation", defined as compassionate, caring, appreciative physical touching (e.g. hugging), and "verbal appreciation", defined as compassionate, caring, appreciative talk that targeted the employee as a person (e.g. "You are so sweet"). Demands

included “physical resistance”, defined as refusing cooperation in caring activity (e.g. not lifting the arm when told to do so),”verbal resistance”, defined as refusing cooperation or showing disagreement with the caring activity (e.g. yelling), “physical aggression”, defined as violent behavior (e.g. hitting), and “verbal aggression”, defined as negative speech targeted towards the employee as a person (e.g. “You are mean”). The emotional reactions were registered as single point events. If the reactions (verbal or physical) continued for a longer period of time with no clear separation between events, the rater made a registration every 10 s.

Part 4 of the observation instrument (item 25 and 26 in Table 1) recorded barriers (i.e. interruptions and impediments) for the eldercare worker in the care of the residents. These measurements were based on Action-Regulation Theory (Hacker, 1994) and their application in an earlier study in Danish eldercare (Jakobsen et al., 2015). An interruption was defined as an event of adequate importance and nature that would significantly interrupt the eldercare worker in performing a work task (e.g. colleague or other resident requesting urgent help). An impediment was defined as an obstacle to fulfill a specific task that required extra work or effort to overcome (e.g. broken or missing assistive device). The rater manually registered the duration during which the eldercare worker was interrupted or hindered in the performance of a task and when the task was continued (“start-stop”). It was also registered what caused the interruption or impediment (coded as descriptive factor 1 in the instrument).

The observation procedure was described in detail to the raters in an observation manual including an appendix with pictures of all available assistive devices for manual handling at nursing homes in Denmark.

2.2. Reliability evaluation

One-hundred and seventeen observations were simultaneously carried out by two raters for the reliability evaluation. These simultaneously performed pairs of observations are referred to as inter-rater observations.

2.2.1. Nursing homes, eldercare workers, residents and raters

This inter-rater evaluation study was conducted at 11 out of the total 20 participating nursing homes in DOSES. On average, 10.6 (SD 7.37) inter-rater observations were conducted at each of the 11 nursing homes. The inter-rater observations were performed on 32 eldercare workers in the care for 75 residents.

Sixteen raters were involved in the inter-rater observations. The raters were either students or graduate students from sports science, physical activity and health science, public health science or occupational therapy.

Because the inter-rater observations were conducted at several time points during the 1.5 year data collection period, the group of raters involved in the inter-rater observations was not constant. Nine raters performed inter-rater observations from September 2013. During spring 2014, four raters stopped and three new raters were trained and performed observations. In addition four new raters were trained and performed inter-rater observations until December 2014.

2.2.2. Rater training

The training of the raters consisted of three sessions of 1–2 h each. The first session provided information about DOSES and the theory behind the observation instrument. The second session included a review of the observation manual and the technical settings of the tablets. Before the third session, the raters used the observation instrument while watching 48 videos of different patient handling situations in laboratory settings and read 11 cases with different verbal and physical interactions between the eldercare worker and a resident. At the third training session, their ratings during the videos and cases were evaluated and discussed. After these training sessions, a minimum of two days field training at nursing homes were completed. By the end of the second day in the field, an experienced rater evaluated if the newly trained rater was sufficiently trained and prepared to perform observations in DOSES. The final evaluation was based on the observational agreement with the experienced rater and an overall judgement of the successful completion of the training.

2.2.3. Data collection

The inter-rater observations were conducted along with the individual observations in DOSES. During the 1.5 years long data collection period in DOSES, the inter-rater observations were performed when logistically possible (depending on number of raters available). It was ensured that inter-rater observations in both day shift and evening shift were represented. In total, 77 inter-rater observations were performed in day shift (66%) and 40 inter-rater observations were performed in evening shifts (34%). Sixty-seven inter-rater observations were conducted at the first 10 nursing homes from September 2013 to May 2014 (time period 1), and 50 inter-rater observations were conducted at the remaining 10 nursing homes from May to December 2014 (time period 2).

During the inter-rater observations, two raters followed the same eldercare worker and performed registrations of the same caring situation while being present in the room at the same time. This procedure controlled for variability in assessment situations. The raters were not permitted to communicate to each other, and they were instructed to position the tablets at an angle that would not allow the other rater to see the registrations performed. Because the inter-rater observations were carried out directly on site, it was not possible to know in advance how many times a situation or activity would occur.

The inter-rater observations varied in length depending on the duration of the total caring activities carried out with a resident. On average, the inter-rater observations, lasted for 13 min (SD 14.1), calculated from entering the room of the resident until leaving the room after the caring activity had been finalized. The 16 raters performed 15 inter-rater observations on average (SD 10.9) with various pairings. Rater 1 and rater 2 position within a pair was assigned randomly.

2.2.4. Inter-rater reliability

The primary purpose of the DOSES observation instrument was to collect information on both the physical and psychosocial work exposures during eldercare work in nursing homes. The inter-rater reliability evaluation was therefore conducted on items 10 to 26 (Table 1) regarding manual handling activities, psychosocial interactions and barriers for carrying out tasks. Items 1 to 9 regarding setting and surroundings were not evaluated.

For a specific item, the inter-rater reliability was evaluated as the agreement of instantaneous occurrence of the item between the two raters.

We found that in the daily work of eldercare workers, many of the defined items in the DOSES observation instrument occurred only rarely (e.g. appreciation, resistance and aggression from resident, support stockings on the resident). In such a case, the agreement between raters would to a large extent reflect the agreement on absence of the risk factor rather than its presence. Therefore, only inter-rater observations with at least one registered occurrence were included in the analysis for that particular item. The number of inter-rater observations excluded from the specific analyses of each item due to no occurrence of that particular item is shown in Appendix A.

Because of the low occurrence of several items, some items were merged for the inter-rater analyzes. Items 10–15 (Table 1) regarding manual handling activities were merged into “Lifting of the resident” (item 10 and 11), “Repositioning of the resident” (item 12 and 13) and “Turning of the resident” (item 14 and 15). Items regarding the psychosocial interactions, were merged into “Appreciation from resident” (item 19 and 20), “Resistance from resident” (item 21 and 22) and, “Aggression from resident” (item 23 and 24). Further, we created a new item “Negative behavior” that combined “resistance“ and “aggression“

(items 21 to 24).

2.3. Data analyses

2.3.1. Data management

During data collection, the tablets were regularly synchronized to the same main computer. Thus, providing the same absolute time to the tablets and hereby a synchronized time-string of the registrations of events. The absolute time was used when comparing two inter-rater observation sequences.

Five out of the total 117 inter-rater observations showed, after visual inspection and data analysis, technical errors due to lack of synchronization. These five inter-rater observations were synchronized by the difference of means between the two sets of observations.

One inter-rater observation showed, after visual inspection and examination by box plot, substantial outlying disagreement in occurrence of events. The inter-rater observation was not found trustworthy (likely due to misunderstanding or a technical error), and was excluded from the data set, resulting in a total number of 116 inter-rater observations for the statistical analyses.

For evaluating the inter-rater reliability, two analytical models were chosen (Fig. 1). For both models, the registrations from the two raters were time stamped and compared within a certain time window. For the manual handling activities, interruptions and impediments (Model 1a, Fig. 1), the time window was calculated separately for every item in each respective inter-rater observation. The duration of the time window was set to avoid overlap of events and to take into account a potential lack of synchronization of events between raters. Thus, the time window encircling every event of an item was calculated as the minimum time difference (Δt) between registrations for the given item. Δt was set to a maximum of 60 s.

Because of the nature of the psychosocial interactions (i.e. appreciation, resistance and aggression from the resident), where the events often occur as flow of verbal or physical outbursts, Δt was set to 60 s to avoid very small Δt and bias towards low agreement (Model 1b, Fig. 1). If both raters registered one or more occurrences of events within the same time window, it would count as an agreement.

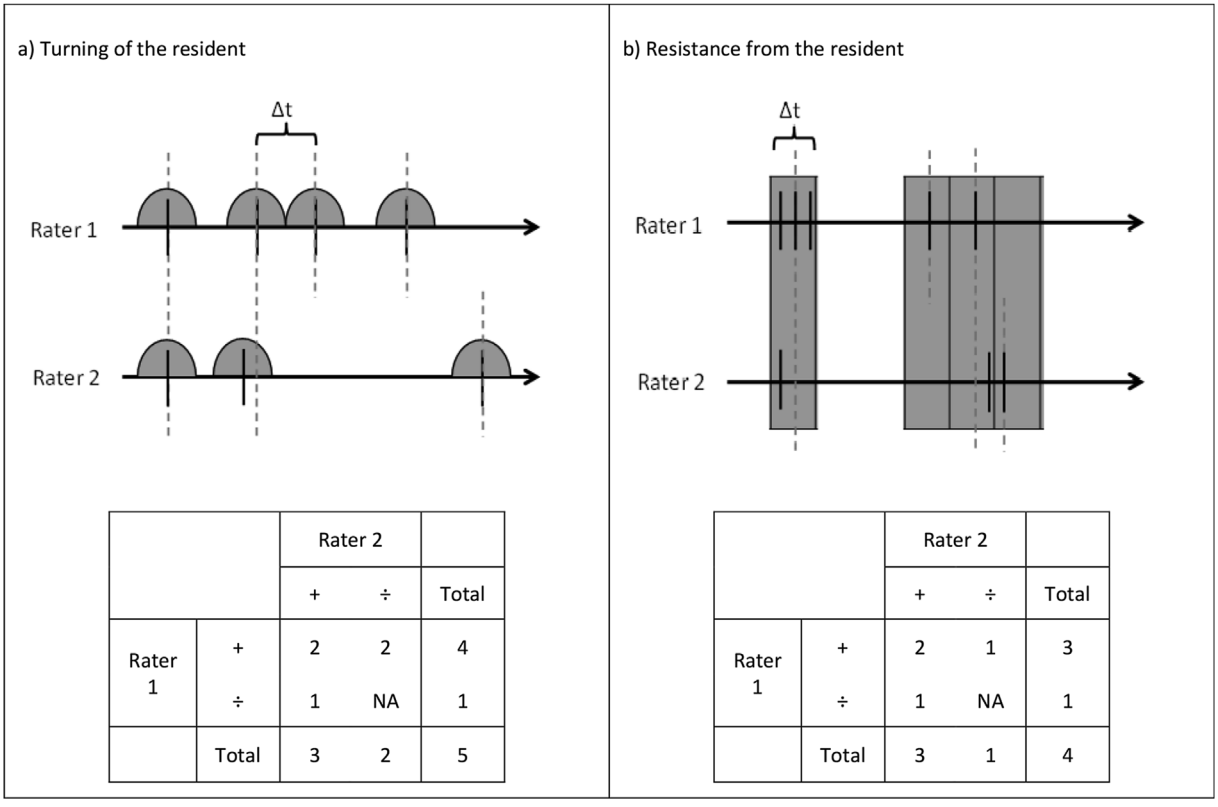
MATLAB (The MathWorks Inc., Natick, MA, USA) was used for the data analyses.

2.3.2. Statistical analyses

Agreement between raters was evaluated using percent agreement as well as the agreement coefficient (AC1) with 95% CI (Gwet, 2014). Percent agreement was calculated as the total number of registrations in which the raters recorded the same item divided by the total number of registrations where at least one rater had recorded the particular item.

The Kappa coefficient (Cohen, 1960) is commonly used for inter-rater agreement calculations. However, it is well known that in situations with rare observed events, Kappa often yields very low coefficients compared to the percent agreement (Cicchetti and Feinstein, 1990; Gwet, 2014). Furthermore, all four marginal totals in the contingency table are required for the calculation of Kappa.

As a result of the analytical models chosen for this study (illustrated



		Rater 2		
		+	÷	Total
Rater 1	+	2	2	4
	÷	1	NA	1
	Total	3	2	5

		Rater 2		
		+	÷	Total
Rater 1	+	2	1	3
	÷	1	NA	1
	Total	3	1	4

Fig. 1. Illustration of a fictive inter-rater observation and corresponding contingency tables of two different items (i.e. "Turning the resident", "Resistance from the resident") representing the two analytical models used in this study.
The model shown in Fig. 1a was used for the manual handling activities, interruptions and impediments, whereas the model shown in Fig. 1b was used for the psychosocial interactions. In model 1a, Δt is calculated as the minimum time difference between registrations. In model 1b, Δt was set fixed to 60 s. Registrations from rater 1 and 2 were time stamped and compared within a certain time window. The size of the time window encircling every event was set to the size of Δt.
Horizontal arrow = time.
Vertical solid-line = registered events.
Colored area = time window.
Broken line across rater 1 and 2 = both raters have registered occurrence of event (agreement on presence)
Broken line only across rater 1 or 2 = only one rater have registered occurrence of event (disagreement on presence).

in Fig. 1), data on agreement of no occurrence of events were not available (the bottom right cell of marginal totals). This data could be estimated but would be severely influenced by assumptions. Furthermore, if data on agreement on no occurrence of events were estimated, it would result in very unbalanced marginal totals because of the rare occurrence. Therefore, we found the Kappa coefficient not suitable for inter-rater agreement calculations in this study. Instead, we used Gwet's AC1 coefficient for the statistical analyses. Gwet's AC1 coefficient provides an alternative to the coefficient calculation for the agreement by chance and is more resistant to unbalanced data (Gwet, 2014). The calculation of AC1 does not include data from the bottom right corner cell of the contingency Table. AC1 can range from −1 to 1. A commonly used three-degree scale was used for interpretation of the agreement coefficient (AC1): poor agreement for AC1 0–0.40; fair to good agreement for AC1 0.40–0.75; and excellent agreement for AC1 > 0.75 (Fleiss et al., 2003). AC1 < 0 was interpreted as no agreement. This indicates an absence of agreement among raters beyond chance.

The statistical calculations for the inter-rater reliability were carried out using scripts written in MATLAB (The MathWorks Inc., Natick, MA, USA).

Differences between inter-rater reliability at time period 1 (T1) and time period 2 (T2) were analyzed using χ^2 tests.

As recommended (Kottner et al., 2011), we tested if the observations for the inter-rater agreement evaluation were representative regarding occurrence of events in daily eldercare work. No statistical differences were found regarding the frequencies of occurrence of events in daily eldercare work between the two sets of data conducted by rater 1 and rater 2 which constitutes the inter-rater observations. Therefore, we compared the observations from this study conducted by rater 1 (n = 116) with the individual observations conducted in DOSES (n = 4600). An independent sample *t*-test was used for analyzing potential differences between the two different sets of observational data. Data were analyzed using SPSS version 21 (SPSS Inc., Chicago, IL).

2.4. Ethical considerations

The study was approved by the Danish Data Protection Agency and the Ethics Committee for the regional capital of Denmark (H-4-2013-028).

3. Results

3.1. Inter-rater reliability of the DOSES observation instrument and differences between exposure types

Overall, the percent agreement ranged from 30.8% to 100% between the items (Table 2) with corresponding agreement coefficients (AC1) ranging from −0.27 to 1.0.

The agreement coefficients (AC1) for the manual handling activities where higher than 0.75 for “Lifting of the resident”, “Turning of the resident”, “Support stockings on resident” and “Squatting”, and therefore designated as “Excellent”. “Push/pull resident” had a “Fair to good” agreement with an agreement coefficient (AC1) of 0.70 and “Repositioning of the resident” had a “Poor” agreement with an agreement coefficient of 0.31.

”Appreciation” (AC1: −0.27) and “Aggression” from the resident (AC1: −0.12) were designated with “No agreement” whereas “Resistance” from the resident (AC1: 0.42) had a “Fair to good” agreement. Combining resistance and aggression into the new variable “Negative behavior” resulted in an AC1 of 0.70 and the designation of “Fair too good” agreement.

Interruptions and impediments showed agreement coefficients (AC1) of 0.70 (“Fair to good“) and 0.88 (“Excellent“), respectively.

Regarding the reliability of the specifics of manual handling activities (i.e. help from resident, support from colleague, use of assistive device), agreement on “No support from colleague” were lower compared to the other specifics (i.e. Use/no use of assistive device, support/no support from resident and support from colleague).

3.2. Development of inter-rater reliability over time of data collection

No significant differences regarding inter-rater reliability were found between the first time period T1 (first 10 nursing homes in the time period from September 2013 to May 2014) and the second time period T2 (last 10 nursing homes in the time period from May to December 2014) (see Appendix B).

3.3. Representativeness of the observations

We found no differences in the average duration of the inter-rater observations (n = 116) and the individual DOSES observations (n = 4600) (Table 3). We did neither find differences between the observations regarding average number of lifting and turning of the resident, aggression and resistance from the resident, interruptions and impediments. Small differences were found between the inter-rater observations and the individual DOSES observations for repositioning of the resident, push/pull of the resident in portable chair and appreciation from the resident.

4. Discussion

The inter-rater reliability of the DOSES observation instrument was generally good. Fourteen of the 18 items showed fair to excellent agreement between raters. The inter-rater agreement did not change over the long data collection period, and the inter-rater observations for the analyses of reliability were predominantly representative regarding occurrence of events in eldercare work.

4.1. Inter-rater reliability of the DOSES observation instrument

Twelve of the items from the DOSES observation instrument had a percent agreement between raters higher than 60%, and level of agreement designated as fair to excellent. Overall, this supports that the DOSES observation instrument is suitable for assessing physical and psychosocial risk factors for MSD in elder care.

The reliability coefficients were most likely conservative estimates for several reasons. We used real-time observations conducted during the actual delivery of work which included several sources of potential bias as compared to laboratory-setting observations or post-hoc video observations. This method is likely to have reduced the reliability, but increased its external validity. Furthermore we based our reliability estimates not on the agreement of non-occurrence of events but solely on the agreement of the occurrence of an event. We only included inter-rater observations with at least one registration of a particular item to not spuriously inflate the concordances by agreement of non-occurrence of rare events. Finally the time window used for determining an agreement or disagreement between raters was set to be short.

The reliability of the DOSES observation instrument cannot be directly compared with the reliability of other observation instruments assessing physical and psychosocial risk factors for MSD in eldercare work in nursing homes. This is primarily because other instruments have different aims (e.g. safe transfer technique), a different focus (e.g. on postures, movements, specific body areas), or use other methods for performing the observation (e.g. video-based, diary, checklist) (Johnsson et al., 2004; Kjellberg et al., 2000; Warming et al., 2004).

4.2. Differences in reliability between exposure types

The inter-rater agreement ranged from 55 to 100% for the items concerning manual handling activities, 77–89% for the items interruptions and impediments, and 31–60% for the items concerning

Table 2
Inter-rater reliability of physical and psychosocial risk factors for musculoskeletal disorders in the DOSES observation instrument, expressed as the percentage of agreement (%) between raters (Rater 1 and 2) and agreement coefficient (AC1) with 95% CI and level of agreement. Lifting, repositioning and turning of the resident, psychosocial interactions and specifics of resident handling activities are generated from several of the items in the DOSES observation instrument.

Manual handling activities, psychosocial interactions, interruptions and impediments					Inter-rater reliability			
	Item number (see Table 1)	Rater 1 (n)[a]	Rater 2 (n)[a]	Inter-rater observations (n)[b]	% agreement	AC1[c]	CI95%	Level of agreement[d]
Lifting of the resident	6 + 7	40	38	29	90.2	0.89	0.79;0.99	Excellent
Repositioning of the resident	8 + 9	23	25	22	54.8	0.31	0.03;0.58	Poor
Turning of the resident	10 + 11	52	49	22	90.6	0.90	0.81;0.98	Excellent
Support stockings on resident	12	8	8	4	100.0	1.00	1.00;1.00	Excellent
Push/pull resident	15	65	67	36	76.0	0.70	0.57;0.82	Fair to good
Squatting	18	79	81	25	79.8	0.75	0.65;0.86	Excellent
Appreciation from resident	23 + 24	11	6	9	30.8	−0.27	−0.74;0.21	No agreement
Resistance from resident	21 + 22	36	49	8	60.4	0.42	0.22;0.61	Fair to good
Aggression from resident	19 + 20	21	9	3	36.4	−0.12	−0.49;0.24	No agreement
Negative behavior[e]	19–22	50	54	9	76.3	0.70	0.56;0.84	Fair to good
Interruptions	26	45	45	28	76.5	0.70	0.56;0.85	Fair to good
Impediments	25	24	27	21	88.9	0.88	0.74;1.00	Excellent
Specifics of resident handling activities								
Use of assistive device	6-11 DF1[f] a-c/d	94	94	38	70.9	0.61	0.50;0.73	Fair to good
No use of assistive device	6-11 DF1[f] d/e	21	18	14	77.3	0.72	0.49;0.94	Fair to good
Support from resident	6 + 8+10	53	50	37	80.7	0.77	0.64;0.89	Excellent
No support from resident	7 + 9+11	62	62	26	67.6	0.56	0.41;0.70	Fair to good
Support from colleague	6-11 DF2[f] a	60	58	19	90.3	0.89	0.81;0.98	Excellent
No support from colleague	6-11 DF2[f] b	55	54	32	51.4	0.23	0.05;0.42	Poor

[a] The total number of occurrences registered by each rater.
[b] The total number of inter-rater observations where the item is registered.
[c] Gwet's AC1 Coefficient (Gwet, 2014).
[d] Classification by Fleiss (Fleiss et al., 2003), AC1 ≤ 0 is classified as no agreement.
[e] Negative behavior = resistance + aggression from resident.
[f] DF1 (Descriptive factor 1) and DF2 (Descriptive factor 2) serve as subcategories of the item (see Table 1).

Table 3
Duration, number of manual handling activities, psychosocial interactions, interruptions and impediments registered per observation in the present study evaluating the inter-rater reliability of the DOSES observation instrument (Rater 1) (n = 116) and in a larger cohort study "Danish observational study of eldercare work and musculoskeletal disorders" (DOSES) using the same instrument for observations in eldercare (n = 4600).

	Inter-rater observations (Rater 1) (N = 116)		DOSES observations (N = 4600)		
	Mean	SD	Mean	SD	Level of sign.[a]
Duration (min)	13.03	14.11	11.72	13.80	0.31
Lifting of the resident	0.35	0.71	0.24	0.60	0.10
Repositioning of the resident	0.18	0.49	0.27	0.79	< 0.05
Turning of the resident	0.43	1.02	0.26	0.93	0.08
Push/pull resident	0.59	1.15	0.37	0.94	< 0.05
Appreciation from resident	0.11	0.43	0.25	0.98	< 0.01
Resistance from resident	0.53	2.73	0.52	2.89	0.99
Aggression from resident	0.22	2.05	0.08	0.84	0.49
Interruptions	0.41	0.93	0.28	0.78	0.14
Impediments	0.20	0.53	0.21	0.57	0.80

[a] Differences between groups are analyzed using an independent *t*-test.

emotional interactions. The level of agreement was fair to excellent for manual handling activities (except for reposition of the resident, which was poor), interruptions and impediments. Two of the emotional interactions (appreciation and aggression from resident) showed no agreement, whereas resistance from resident showed fair to good agreement. The agreement for manual handling activities were in line with the level of agreement for interruptions and impediments, whereas, emotional interactions had lower agreement when used as single items. Merging "Resistance" and "Aggression" from resident into "Negative behavior" brought the negative emotional demands to the same agreement coefficient as push/pull resident and interruptions.

The variation in inter-rater agreement between the exposure types in this study (e.g. manual handling activities, emotional interactions, interruptions, impediments) may be explained by their very different inherent characteristics. Most of the manual handling activities are characterized by single, isolated events occurring with sufficient time between events. Often using an assistive device or getting help from a colleague occurred during the manual handling, which made the situation recognizable and predictable, and thereby easier to classify correctly, resulting in a good agreement (Park et al., 2009). Gross body movements, like squatting, are also easier to observe than smaller motions (Burt and Punnett, 1999). Lower agreement between raters was found for repositioning of the resident compared to the other manual handling activities. This may be explained by the smaller and more spontaneous movements during repositioning than for the other manual handling activities.

Interruptions and impediments were easily recognizable because the work flow was clearly stopped (e.g. colleague entered the room and asked a question, the assistive device was broken, an alarm was set off) that may explain their good inter-rater agreement.

The emotional interactions (i.e. appreciation, resistance and aggression from the resident) generally had a lower agreement between raters than the items for manual handling activities, interruptions and impediments. For the emotional interactions, the lower agreement may

be explained by their common occurrence as flow of speech or physical outbursts (e.g. hitting, kicking, refusing cooperation), which made registering the time and frequency of occurrence difficult. Also, even though the emotional interactions were defined and described clearly in the observation manual, there was still a larger room for interpretation by the observer, compared to the physical handling activities, interruptions and impediments. The higher agreement achieved when merging the two separate item resistance and aggression into one item of negative behavior supports this interpretation. Therefore, our recommendation for future studies applying the DOSES observation instrument is to use negative behavior as a single item without differentiation between aggression and resistance. Moreover, future studies ought to consider setting up emotional interactions in the observation instrument as start-stop events instead of point events to better capture the common occurrence as a flow of speech and physical outbursts.

4.3. Changes in the inter-rater reliability over the data collection period

Because the reliability of an observation instrument may depend on the experience level of raters (Park et al., 2009; Voskuijl and van Sliedregt, 2002), we investigated if the level of agreement was higher in the second than in the first half of the 1.5 year data collection period. However, no differences in inter-rater reliability were found between the first and second half of the 1.5 year data collection period. This might be due to the turnover of raters during the data collection period, where the group of raters as a whole did not get more experienced throughout the data collection period. Another explanation could be that the comprehensive training of the raters may have brought them up to a high sustained level of skill. Because of the design of this study and the limited amount of data, we could not investigate whether raters who were present during the whole data collection improved their rater skills. We consider the similar level of inter-rater agreement throughout a long data collection period as a strength of the observation instrument. It is recommended not to neglect the importance of thorough training of raters in observation studies, as it could prevent differences in data quality as result of increasing experience level through the data collection period.

4.4. Representability of the data used for evaluating the inter-rater reliability of the observation instrument

That two raters instead of one follow a worker may influence the behaviors of both the resident and the worker. Moreover, permission from the workers to carry out the inter-rater observations at vulnerable residents could also depend on the number of observers. Thus, the observations collected for evaluating the inter-rater reliability could differ from the observations in the larger DOSES study. However, we found no differences of substantial size between the inter-rater observations (n = 116) and the more than 4000 observations performed in the entire DOSES study, suggesting that the inter-rater evaluation in this study is representative for occurrence of events in daily eldercare work.

4.5. The DOSES observation instrument and its applicability for studying risk factors for MSD in eldercare work

Manual handling activities have been shown to be associated with MSD and its consequences among nurses and eldercare workers (Lagerström et al., 1998). Many observation methods have focused on ergonomic body positions, movements, load on specific body parts and safe transfer technique (Johnsson et al., 2004; Kjellberg et al., 2000; Takala et al., 2010; Warming et al., 2004). Today, technical methods (e.g. accelerometers) have been developed, so physical activity types (e.g. walking, running, cycling), body postures (e.g. sitting, standing) and movements (e.g. forward bending of the back and arm elevation) (Korshøj et al., 2014; Skotte et al., 2014; Stemland et al., 2015) can be assessed in larger study populations at relatively low cost. Therefore, the items developed for this observation instrument regarding physical risk factors for MSD focused on the manual handling tasks and the external load which cannot be attained from the technical methods. For example, in DOSES, the activity types and body positions were measured by accelerometers while the information of manual handling and use of equipment were collected by workplace observations.

To the best of our knowledge, there are no epidemiological studies that have examined observer-assessed psychosocial working conditions and risk of MSD among eldercare workers. In a study of transit operators, Greiner and Krause (2006) reported an association of observer-assessed barriers at work with a higher level of MSD. Among eldercare workers, Jakobsen et al. (2015, 2016) found that observer-assessed low regulation requirements (Jakobsen et al., 2015) and high amounts of social interactions between eldercare workers and residents (Jakobsen et al., 2016) were associated with a higher level of depressive symptoms among the eldercare workers. The instrument used by Jakobsen et al. in these two studies was the basis for the development of the psychosocial part of the DOSES observational instrument.

The validity of the DOSES observational instrument has not been evaluated. However, for achieving best possible validity, we undertook 4 activities: (1) the instrument was based on previously used and well-functioning observation instruments (Jakobsen et al., 2016, 2015; Koppelaar et al., 2012), (2) experienced designers of previous observational instruments were actively involved in the development process, (3) a thorough manual was written clearly describing every item's definition and rating, and (4) items were excluded from the final observational instrument where high agreement could not be reached.

The equipment used in performing the DOSES observation instrument was a tablet equipped with specific software to systematically collect data. It is our experience that the instrument can be used by everyone regardless of previous experience of eldercare work, but rigorous training in observation is needed. The observation instrument can be used in different settings (e.g. the resident's room, at the ward, outside nursing home) where interaction between eldercare worker and resident takes place. It is our impression from conducting observations at 20 nursing homes in the larger DOSES study, using the DOSES observation instrument, that the DOSES observation instrument shows good feasibility. The instrument was designed for use in elder care, but could potentially also be used with some modifications in jobs with similar settings and work tasks (e.g. hospitals, home care).

4.6. Strengths and limitations

The primary strength of this study is that the inter-rater reliability of the instrument was evaluated based on many pairs of observations from 11 representative Danish nursing homes and a wide variety of residents over a long period of time. This allowed us to investigate if the reliability differed over the data collection period and if the inter-rater observations were representative for the occurrence of events in daily eldercare work collected in the larger DOSES study using the same instrument. Moreover, the inter-rater reliability evaluation was conducted in natural settings and not in artificially established situations or in a laboratory. The many participating nursing homes, wide variety of residents, and the natural setting all contributes to increasing the external validity of the results. Another strength is the thorough development of the instrument with a detailed observation manual and the comprehensive training of the raters.

A limitation of this study was that some items rarely occurred (e.g. support stockings on resident, appreciation and aggression from

resident). This provided limited events for analyses and thereby restricted the statistical power. The limited amount of data material also hindered analyses of potential interactions on the inter-rater reliability of physical and psychosocial factors (e.g. if emotional reactions from residents influenced the inter-rater agreement on manual handling). Furthermore, the reliability estimates of the evaluation were limited by the lack of systematic pairings of raters and the variation in the number of observations performed by each rater. The pairing of the raters was not systematically determined, but determined by logistics, e.g. the availability of specific raters at specific times during the data collection period, resulting in a variation in pairings and number of observations performed by each rater. However we do not expect this to have affected the reliability estimates in a major way as most raters were paired with a range of other raters. Moreover, potential differences in the experience of the raters may have influenced the reliability results of the present study towards the null, although, this influence would be small due to extensive training of all raters.

5. Conclusions

The DOSES observation instrument permits simultaneously assessment of physical and psychosocial risk factors for MSD in eldercare work. The inter-rater reliability of the instrument was generally good, though the agreement differed between exposure types. Of the 18 examined items, 14 could be observed with fair to excellent agreement. "Reposition of the resident" and "resistance from the resident" showed poor agreement and two psychosocial items (i.e. appreciation and aggression from the resident) showed no agreement. Combining "aggression from resident" with "resistance from resident" resulted in a new item of "negative behavior" with fair to good agreement. The reliability did not differ between the first and second half of the data collection period, and the observations for the analyses of reliability were overall representative regarding occurrence of events in daily eldercare work. The DOSES observation instrument showed to be feasible in a large project where data was collected by many individuals over a longer time period.

Funding

The study was financed by a grant from the Danish government.

Acknowledgements

The authors wish to thank Anette Fauerskov, Bruus Jørgensen and Rikke Berner Rasmussen for their devoted work with the development of the observation instrument. We also gratefully acknowledge the entire DOSES research group and personnel who contributed to the data collection, especially Dorte Ekner and Rikke Berner Rasmussen, and the 20 nursing homes that participated in DOSES. Finally, thanks to Hans Bay for statistical assistance.

Appendix A

Table A.1
Percentage of the 116 inter-rater observations, without occurrence of the specific item, and with at least one rater registered ≥1 occurrence of the item.

Manual handling activities, psychosocial interactions, interruptions and impediments	Inter-rater observations with no registrations of the specific item		Inter-rater observations with ≥1 registrations of the specific item	
	n	%	n	%
Lifting of the resident	87	75.0	29	25.0
Repositioning of the resident	94	81.0	22	19.0
Turning of the resident	94	81.0	22	19.0
Appreciation from resident	107	92.2	9	7.8
Resistance from resident	108	93.1	8	6.9
Aggression from resident	113	97.4	3	2.6
Support stockings on the resident	112	96.6	4	3.4
Push/pull resident	80	69.0	36	31.0
Squatting	91	78.4	25	21.6
Interruptions	88	75.9	28	24.1
Impediments	95	81.9	21	18.1
Negative behavior[a]	107	92.2	9	7.8
Specifics of resident handling activities				
Use of assistive device	78	67.2	38	32.8
No use of assistive device	102	87.9	14	12.1
Support from resident	79	68.1	37	31.9
No support from resident	90	77.6	26	22.4
Support from colleague	97	83.6	19	16.4
No support from colleague	84	72.4	32	27.6

[a] Negative behavior = resistance + aggression from resident.

Appendix B

Table B.1
Inter-rater reliability of physical and psychosocial risk factors in the DOSES observation instrument stratified by two time periods; T1 (first 10 nursing homes; September 2013 to May 2014) and T2 (last 10 nursing homes; May to December 2014). Inter-rater reliability is expressed as the percentage of agreement (%) between pair of raters (Rater 1 and 2) and agreement coefficient (AC1) with 95% CI and level of agreement.

Manual handling activities, psychosocial interactions, interruptions and impediments				Inter-rater reliability		
	Rater 1 (n)[a]	Rater 2 (n)[a]	Inter-rater observations (n)[b]	% agreement	AC1[c]	CI95%
Lifting of the resident, T1	19	19	16	90.0	0.89	0.74;1.00
Lifting of the resident, T2	21	19	13	90.5	0.90	0.75;1.00
Repositioning of the resident, T1	15	16	14	55.0	0.31	−0.03;0.65
Repositioning of the resident, T2	8	9	8	54.5	0.30	−0.18;0.78
Turning of the resident, T1	29	28	11	90.0	0.89	0.77;1.00
Turning of the resident, T2	23	21	11	91.3	0.91	0.78;1.00
Support stockings on resident, T1	4	4	2	100.0	1.00	1.00;1.00
Support stockings on resident, T2	4	4	2	100.0	1.00	1.00;1.00
Push/pull resident, T1	26	23	17	69.0	0.58	0.35;0.81
Push/pull resident, T2	39	44	19	80.4	0.76	0.62;0.90
Squatting, T1	37	36	9	82.5	0.79	0.65;0.93
Squatting, T2	39	44	16	77.6	0.72	0.57;0.87
Appreciation from resident, T1	4	3	4	40.0	−0.03	−0.86;0.79
Appreciation from resident, T2	7	3	5	25.0	−0.41	−1.00;0.19
Resistance from resident, T1	19	21	5	66.7	0.54	0.27;0.81
Resistance from resident, T2	17	28	3	55.2	0.31	0.03;0.60
Aggression from resident, T1	2	0	1	0.0	−1.00	−1.00;-1.00
Aggression from resident, T2	19	9	2	40.0	−0.03	−0.41;0.35
Interruptions, T1	21	21	14	75.0	0.68	0.45;0.91
Interruptions, T2	24	24	14	77.8	0.72	0.52;0.92
Impediments, T1	9	10	10	90.0	0.89	0.67;1.00
Impediments, T2	15	17	11	88.2	0.87	0.69;1.00
Negative behavior[d], T1	20	21	5	70.8	0.61	0.36;0.86
Negative behavior[d], T2	30	33	4	80.0	0.76	0.59;0.92
Specifics of resident handling activities						
Use of assistive device, T1	61	61	24	69.4	0.59	0.44;0.73
Use of assistive device, T1	33	33	14	73.7	0.66	0.48;0.84
No use of assistive device, T1	2	2	2	100.0	1.00	1.00;1.00
No use of assistive device, T2	19	16	12	75.0	0.68	0.43;0.93
Support from resident, T1	29	26	22	77.4	0.72	0.53;0.91
Support from resident, T2	24	24	15	84.6	0.82	0.66;0.99
No support from resident, T1	34	37	14	65.1	0.51	0.31;0.71
No support from resident, T2	28	25	12	71.0	0.61	0.40;0.83
Support from colleague, T1	31	31	8	87.9	0.86	0.74;0.99
Support from colleague;T2	29	27	11	93.1	0.93	0.83;1.00
No support from colleague, T1	32	32	22	45.5	0.10	−0.15;0.34
No support from colleague, T2	23	22	10	60.7	0.43	0.16;0.70

[a] The total number of occurrences registered by each rater.
[b] The total number of inter-rater observations where the item is registered.
[c] Gwet's AC1 Coefficient (Gwet, 2014).
[d] Negative behavior = resistance + aggression from resident.

References

Aagestad, C., Tyssen, R., Johannessen, H.A., Gravseth, H.M., Tynes, T., Sterud, T., 2014. Psychosocial and organizational risk factors for doctor-certified sick leave: a prospective study of female health and social workers in Norway. BMC Publ. Health 14. http://dx.doi.org/10.1186/1471-2458-14-1016.

Andersen, L.L., Clausen, T., Mortensen, O.S., Burr, H., Holtermann, A., 2012. A prospective cohort study on musculoskeletal risk factors for long-term sickness absence among healthcare workers in eldercare. Int. Arch. Occup. Environ. Health 85, 615–622. http://dx.doi.org/10.1007/s00420-011-0709-5.

Buchholz, B., Paquet, V., Punnett, L., Lee, D., Moir, S., 1996. PATH: a work sampling-based approach to ergonomic job analysis for construction and other non-repetitive work. Appl. Ergon. 27, 177–187. http://dx.doi.org/10.1016/0003-6870(95)00078-X.

Burt, S., Punnett, L., 1999. Evaluation of interrater reliability for posture observations in a field study. Appl. Ergon. 30, 121–135.

Cicchetti, D.V., Feinstein, A.R., 1990. High agreement but low kappa: II. Resolving the paradoxes. J. Clin. Epidemiol. 43, 551–558.

Clausen, T., Andersen, L.L., Holtermann, A., Jorgensen, A.F.B., Aust, B., Rugulies, R., 2013. Do self-reported psychosocial working conditions predict low back pain after adjustment for both physical work load and depressive symptoms? A prospective study among female eldercare workers. Occup. Environ. Med. 70, 538–544. http://dx.doi.org/10.1136/oemed-2012-101281.

Cohen, J., 1960. A coefficient of agreement for nominal scales. Educ. Psychol. Meas. 20, 37–46. http://dx.doi.org/10.1177/001316446002000104.

Davis, K.G., Kotowski, S.E., 2015. Prevalence of musculoskeletal disorders for nurses in

hospitals, long-term care facilities, and home health care: a comprehensive review. Hum. Factors J. Hum. Factors Ergon. Soc. 57, 754–792. http://dx.doi.org/10.1177/0018720815581933.

del Campo, M.T., Romo, P.E., de la Hoz, R.E., Villamor, J.M., Mahíllo-Fernández, I., 2017. Anxiety and depression predict musculoskeletal disorders in health care workers. Arch. Environ. Occup. Health 72, 39–44. http://dx.doi.org/10.1080/19338244.2016.1154002.

Fleiss, J.L., Levin, B., Paik, M.C., 2003. Statistical methods for rates and proportions. In: Wiley Series in Probability and Statistics, third ed. J. Wiley, Hoboken, N.J.

Greiner, B.A., Krause, N., 2006. Observational stress factors and musculoskeletal disorders in urban transit operators. J. Occup. Health Psychol. 11, 38–51. http://dx.doi.org/10.1037/1076-8998.11.1.38.

Gupta, N., Heiden, M., Mathiassen, S.E., Holtermann, A., 2016. Prediction of objectively measured physical activity and sedentariness among blue-collar work using survey questionnaires. Scand. J. Work. Environ. Health. http://dx.doi.org/10.5271/sjweh.3561.

Gwet, K.L., 2014. Handbook of inter-rater reliability: the definitive guide to measuring the extent of agreement among raters; [a handbook for researchers, practitioners, teachers & students]. In: Advanced Analytics, 4. ed. LLC, Gaithersburg, MD.

Hacker, W., 1994. Action regulation theory and occupational psychology: review of German empirical research since 1987. Ger. J. Psychol. 18, 91–120.

Hauke, A., Flintrop, J., Brun, E., Rugulies, R., 2011. The impact of work-related psychosocial stressors on the onset of musculoskeletal disorders in specific body regions: a review and meta-analysis of 54 longitudinal studies. Work. Stress 25, 243–256. http://dx.doi.org/10.1080/02678373.2011.614069.

Jakobsen, L.M., Jorgensen, A.F.B., Thomsen, B.L., Albertsen, K., Greiner, B.A., Rugulies, R., 2016. Emotion work within eldercare and depressive symptoms: a cross-sectional multi-level study assessing the association between externally observed emotion work and self-reported depressive symptoms among Danish eldercare workers. Int. J. Nurs. Stud. 62, 183–192. http://dx.doi.org/10.1016/j.ijnurstu.2016.07.021.

Jakobsen, L.M., Jorgensen, A.F.B., Thomsen, B.L., Greiner, B.A., Rugulies, R., 2015. A multilevel study on the association of observer-assessed working conditions with depressive symptoms among female eldercare workers from 56 work units in 10 care homes in Denmark. BMJ Open 5http://dx.doi.org/10.1136/bmjopen-2015-008713.e008713–e008713.

Jensen, L.D., Ryom, P.K., Christensen, M.V., Andersen, J.H., 2012. Differences in risk factors for voluntary early retirement and disability pension: a 15-year follow-up in a cohort of nurses' aides. BMJ Open 2, e000991. http://dx.doi.org/10.1136/bmjopen-2012-000991.

Johnsson, C., Kjellberg, K., Kjellberg, A., Lagerström, M., 2004. A direct observation instrument for assessment of nurses' patient transfer technique (DINO). Appl. Ergon. 35, 591–601. http://dx.doi.org/10.1016/j.apergo.2004.06.004.

Kjellberg, K., Johnsson, C., Proper, K., Olsson, Elisabeth, Hagberg, M., 2000. An observation instrument for assessment of work technique in patient transfer tasks. Appl. Ergon. 31, 139–150. http://dx.doi.org/10.1016/S0003-6870(99)00046-0.

Koch, M., Lunde, L.-K., Gjulem, T., Knardahl, S., Veiersted, K.B., 2016. Validity of questionnaire and representativeness of objective methods for measurements of mechanical exposures in construction and health care work. PLoS One 11, e0162881. http://dx.doi.org/10.1371/journal.pone.0162881.

Koppelaar, E., Knibbe, H.J.J., Miedema, H.S., Burdorf, A., 2012. The influence of ergonomic devices on mechanical load during patient handling activities in nursing homes. Ann. Occup. Hyg. 56, 708–718. http://dx.doi.org/10.1093/annhyg/mes009.

Korshøj, M., Skotte, J.H., Christiansen, C.S., Mortensen, P., Kristiansen, J., Hanisch, C., Ingebrigtsen, J., Holtermann, A., 2014. Validity of the Acti4 software using ActiGraph GT3X+accelerometer for recording of arm and upper body inclination in simulated work tasks. Ergonomics 57, 247–253. http://dx.doi.org/10.1080/00140139.2013.869358.

Kottner, J., Audigé, L., Brorson, S., Donner, A., Gajewski, B.J., Hróbjartsson, A., Roberts, C., Shoukri, M., Streiner, D.L., 2011. Guidelines for reporting reliability and agreement studies (GRRAS) were proposed. J. Clin. Epidemiol. 64, 96–106. http://dx.doi.org/10.1016/j.jclinepi.2010.03.002.

Kwak, L., Proper, K.I., Hagströmer, M., Sjöström, M., 2011. The repeatability and validity of questionnaires assessing occupational physical activity–a systematic review. Scand. J. Work. Environ. Health 37, 6–29.

Lagerström, M., Hansson, T., Hagberg, M., 1998. Work-related low-back problems in nursing. Scand. J. Work. Environ. Health 24, 449–464.

Luime, J.J., Kuiper, J.I., Koes, B.W., Verhaar, J.A.N., Miedema, H.S., Burdorf, A., 2004. Work-related risk factors for the incidence and recurrence of shoulder and neck complaints among nursing-home and elderly-care workers. Scand. J. Work. Environ. Health 30, 279–286.

Miranda, H., Punnett, L., Gore, R., Boyer, J., 2011. Violence at the workplace increases the risk of musculoskeletal pain among nursing home workers. Occup. Environ. Med. 68, 52–57. http://dx.doi.org/10.1136/oem.2009.051474.

Miranda, H., Punnett, L., Gore, R.J., ProCare research team, 2014. Musculoskeletal pain and reported workplace assault: a prospective study of clinical staff in nursing homes. Hum. Factors 56, 215–227.

Park, J.-K., Boyer, J., Tessler, J., Casey, J., Schemm, L., Gore, R., Punnett, L., Promoting healthy and safe employme, 2009. Inter-rater reliability of PATH observations for assessment of ergonomic risk factors in hospital work. Ergonomics 52, 820–829. http://dx.doi.org/10.1080/00140130802641585.

Prince, S.A., Adamo, K.B., Hamel, M., Hardt, J., Connor Gorber, S., Tremblay, M., 2008. A comparison of direct versus self-report measures for assessing physical activity in adults: a systematic review. Int. J. Behav. Nutr. Phys. Activ. 5, 56. http://dx.doi.org/10.1186/1479-5868-5-56.

Skotte, J., Korshøj, M., Kristiansen, J., Hanisch, C., Holtermann, A., 2014. Detection of physical activity types using triaxial accelerometers. J. Phys. Activ. Health 11, 76–84. http://dx.doi.org/10.1123/jpah.2011-0347.

Stemland, I., Ingebrigtsen, J., Christiansen, C.S., Jensen, B.R., Hanisch, C., Skotte, J., Holtermann, A., 2015. Validity of the Acti4 method for detection of physical activity types in free-living settings: comparison with video analysis. Ergonomics 58, 953–965. http://dx.doi.org/10.1080/00140139.2014.998724.

Takala, E.-P., Pehkonen, I., Forsman, M., Hansson, G.-A., Mathiassen, S.E., Neumann, W.P., Sjøgaard, G., Veiersted, K.B., Westgaard, R.H., Winkel, J., 2010. Systematic evaluation of observational methods assessing biomechanical exposures at work. Scand. J. Work. Environ. Health 36, 3–24.

Trinkoff, A.M., Lipscomb, J.A., Geiger-Brown, J., Storr, C.L., Brady, B.A., 2003. Perceived physical demands and reported musculoskeletal problems in registered nurses. Am. J. Prev. Med. 24, 270–275.

Voskuijl, O.F., van Sliedregt, T., 2002. Determinants of interrater reliability of job analysis: a meta-analysis. Eur. J. Psychol. Assess. 18, 52–62. http://dx.doi.org/10.1027//1015-5759.18.1.52.

Warming, S., Juul-Kristensen, B., Ebbehøj, N.E., Schibye, B., 2004. An observation instrument for the description and evaluation of patient transfer technique. Appl. Ergon. 35, 603–614. http://dx.doi.org/10.1016/j.apergo.2004.06.007.

Zapf, D., 2002. Emotion work and psychological well-being. Hum. Resour. Manag. Rev. 12, 237–268. http://dx.doi.org/10.1016/S1053-4822(02)00048-7.