Taylor & Francis
Taylor & Francis Group

Prehospital Emergency Care

ISSN: 1090-3127 (Print) 1545-0066 (Online) Journal homepage: http://tandfonline.com/loi/ipec20

Reliability and Validity of Survey Instruments to Measure Work-Related Fatigue in the Emergency Medical Services Setting: A Systematic Review

P. Daniel Patterson, Matthew D. Weaver, Anthony Fabio, Ellen M. Teasley, Megan L. Renn, Brett R. Curtis, Margaret E. Matthews, Andrew J. Kroemer, Xiaoshuang Xun, Zhadyra Bizhanova, Patricia M. Weiss, Denisse J. Sequeira, Patrick J. Coppler, Eddy S. Lang & J. Stephen Higgins

To cite this article: P. Daniel Patterson, Matthew D. Weaver, Anthony Fabio, Ellen M. Teasley, Megan L. Renn, Brett R. Curtis, Margaret E. Matthews, Andrew J. Kroemer, Xiaoshuang Xun, Zhadyra Bizhanova, Patricia M. Weiss, Denisse J. Sequeira, Patrick J. Coppler, Eddy S. Lang & J. Stephen Higgins (2018) Reliability and Validity of Survey Instruments to Measure Work-Related Fatigue in the Emergency Medical Services Setting: A Systematic Review, Prehospital Emergency Care, 22:sup1, 17-27, DOI: 10.1080/10903127.2017.1376134

To link to this article: https://doi.org/10.1080/10903127.2017.1376134

ORIGINAL RESEARCH

RELIABILITY AND VALIDITY OF SURVEY INSTRUMENTS TO MEASURE WORK-RELATED FATIGUE IN THE EMERGENCY MEDICAL SERVICES SETTING: A SYSTEMATIC REVIEW

P. Daniel Patterson, PhD, NRP, Matthew D. Weaver, PhD, EMT-P, Anthony Fabio, PhD, MPH, Ellen M. Teasley, LAT, ATC, Megan L. Renn, BS, Brett R. Curtis, BS, Margaret E. Matthews, BS, Andrew J. Kroemer, BS, Xiaoshuang Xun, BS, Zhadyra Bizhanova, BS, Patricia M. Weiss, MLIS, Denisse J. Sequeira, BS, Patrick J. Coppler, MSPAS, PA-C, Eddy S. Lang, MDCM, CCFP (EM), J. Stephen Higgins, PhD

ABSTRACT

Background: This study sought to systematically search the literature to identify reliable and valid survey instruments for fatigue measurement in the Emergency Medical Services (EMS) occupational setting. **Methods**: A systematic review study design was used and searched six databases, including one website. The research question guiding the search was developed a priori and registered with the PROSPERO database of systematic reviews: "Are there reliable and valid instruments for measuring fatigue among EMS personnel?" (2016:CRD42016040097). The primary outcome of interest was criterion-related validity. Important outcomes of interest included reliability (e.g., internal consistency), and indicators of sensitivity and specificity. Members of the research team independently screened records from the databases. Full-text articles were evaluated by adapting the Bolster and Rourke system for categorizing findings of systematic reviews, and the rated data abstracted from the body of literature as favorable, unfavorable, mixed/inconclusive, or no impact. The Grading of Recommendations, Assessment, Development and Evaluation (GRADE) methodology was used to evaluate the quality of evidence. **Results**: The search strategy yielded 1,257 unique records. Thirty-four unique experimental and non-experimental studies were determined relevant following full-text review. Nineteen studies reported on the reliability and/or validity of ten different fatigue survey instruments. Eighteen different studies evaluated the reliability and/or validity of four different sleepiness survey instruments. None of the retained studies reported sensitivity or specificity. Evidence quality was rated as very low across all outcomes. **Conclusions**: In this systematic review, limited evidence of the reliability and validity of 14 different survey instruments to assess the fatigue and/or sleepiness status of EMS personnel and related shift worker groups was identified. **Key words:** reliability; validity; fatigue; sleepiness

PREHOSPITAL EMERGENCY CARE 2018;22:17–27

Received July 14, 2017 from the Department of Emergency Medicine, University of Pittsburgh, Pittsburgh, Pennsylvania (PDP, EMT, MLR, BRC, MEM, AJK, XX, ZB, DJS, PJC); Division of Sleep and Circadian Disorders, Departments of Medicine and Neurology, Brigham and Women's Hospital; Division of Sleep Medicine, Harvard Medical School; Boston, Massachusetts (MDW); Department of Epidemiology, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, Pennsylvania (AF); Health Sciences Library System, University of Pittsburgh, Pittsburgh, Pennsylvania (PMW); Emergency Medicine, University of Calgary, Calgary, Alberta, Canada (ESL); National Highway Traffic Safety Administration, Washington, District of Columbia (JSH). Revision received August 18, 2017; accepted for publication August 28, 2017.

Supplemental data for this article can be accessed on the publisher's website.

Work performed on this manuscript was supported with funding from the U.S. Department of Transportation, National Highway Traffic Safety Administration to the National Association of State EMS Officials (NASEMSO): contract/grant number: DTNH2215R00029. The views contained in this article are those of the authors and not necessarily those of the National Highway Traffic Safety Administration.

Address correspondence to P. Daniel Patterson, PhD, University of Pittsburgh, Emergency Medicine, 3600 Forbes Avenue, Iroquois Bldg., Suite 400A, Pittsburgh, PA 15260, USA. Email: pattersonpd@upmc.edu

doi: 10.1080/10903127.2017.1376134

BACKGROUND

Fatigue is a subjective, unpleasant physical and cognitive state, with feelings of tiredness and exhaustion, all contributing to an unrelenting overall condition that impacts the ability to function safely and efficiently (1). Emergency Medical Services (EMS) personnel report high levels of mental and physical fatigue while at work (2, 3). The measurement of fatigue with reliable and valid instruments is a key component of fatigue mitigation; what cannot be reliably measured cannot be addressed (4). There is no agreed upon standard definition of fatigue and no gold standard survey instrument to measure fatigue among EMS or related shift worker groups (1, 5–7). The lack of a standard presents a challenge to EMS administrators and researchers who seek to assess, monitor, and reduce fatigue as part of a robust fatigue risk management program.

Fatigue is typically assessed with a combination of objective and subjective measurement tools (5). Objective assessment has involved measurement of indicators of human physiology or indicators of performance (8). Subjective assessment with survey instruments (questionnaires) is a commonly used technique across populations, and continues to be the principal approach to fatigue measurement for working adults (5). Dozens of survey instruments are potentially capable of measuring one or more dimensions of fatigue (9). Many fatigue survey instruments have been developed for use with a specific population (i.e., patients with cancer) (10, 11). These tools are not well suited for use in working populations like EMS personnel due to specialized item wording, content, and the context in which various instruments were designed. There are also differences in the underlying phenomena the tool is designed to capture; for example, cancer-related fatigue is distinct from work-related fatigue (6). Few fatigue survey instruments have been developed for use in working populations and tested for their reliability and validity (5). For the EMS administrator, use of instruments developed and tested with EMS personnel would, if easy to administer, be ideal for the assessment and monitoring of fatigue in the EMS work environment (2).

This study sought to systematically review the literature and identify reliable and valid survey instruments for fatigue measurement in the EMS occupational setting guided by the single question: "Are there reliable and valid instruments for measuring fatigue among EMS personnel?" (PROSPERO 2016:CRD42016040097) (12).

METHODS

A systematic search of six databases was conducted, including one website (PubMed/Medline, Cumulative Index to Nursing and Allied Health Literature (CINAHL), Scopus, PsycINFO, the Published International Literature on Traumatic Stress (PILOTS) database, and the National Institute of Justice (NIJ) publications and multimedia website). The search was guided by explicit search criteria. The details of the methodology, study protocol, and procedures for adjudicating published and unpublished literature are outlined in a separate publication (13). The elements of the search and protocol unique to this systematic review are described in the following sections.

Study Design

This study included literature that described use of randomized controlled trials, quasi-experimental studies (e.g., before and after designs) (14), and observational study designs (e.g., prospective cohort, cross-sectional, and analyses of secondary or administrative datasets).

Types of Participants

This research included studies that administered occupational fatigue survey instruments to persons 18 years of age and older who were classified as shift workers, EMS personnel, or members of similar worker groups (12).

Types of Interventions

The search was designed to capture studies reporting use of fatigue survey instruments to assess/diagnose fatigue in the EMS workplace or a workplace environment of related shift worker groups. The search was not intended to assess biometric tracking devices, wearable devices, or related instruments that may be portrayed as objective measures of fatigue. Additionally, the purpose of the study did not extend to the evaluation of sleepiness. This research did not specifically search for sleepiness instruments, but rather, focused on occupational fatigue survey instruments. Sleepiness and fatigue are related, but are considered to represent different constructs (7). Sleepiness is described as "one's tendency to fall asleep" (sleep propensity) (15). The Multiple Sleep Latency Test (MSLT) is accepted as a gold standard measure of sleepiness (7). Some degree of sleepiness is common and a normal experience by most individuals over a 24-hour period, regardless of shiftwork (7). Therefore, this research reported on sleepiness instruments tested in EMS shift workers and related worker groups that were captured incidentally even though the search was not designed to capture all available instruments of this type.

Types of Outcome Measures

The GRADE methodology requires the selection of outcomes of interest and rating of those outcomes as

critical and important for purposes of decision-making by the target population and stakeholders (16). An expert panel selected outcomes for this systematic review a priori (12). The primary (critical) outcome of interest was criterion-related validity (predictive validity). Secondary (important) outcomes of interest include reliability (e.g., measures of internal consistency such as Cronbach's alpha and test-retest reliability), indicators of sensitivity, and indicators of specificity.

Search Methods for Studies

The research librarian (PMW) executed searches of five bibliographic database products and one website: PubMed/Medline, the Cumulative Index to Nursing and Allied Health Literature (CINAHL), Scopus, PsycINFO, the Published International Literature on Traumatic Stress (PILOTS), and the publications section of the National Institute of Justice (NIJ) website. See the methods paper for a detailed description of the sources searched and search terms used (13). All searches included literature from January 1980 to October 2016. See Online Supplement Appendix A for search strategy details specific to this systematic review.

Data Collection and Analyses

Two investigators (AF and PJC) independently screened titles and abstracts initially. The Kappa statistic was used to estimate inter-rater agreement.

Full-Text Review

Seven different co-investigators (EMT, MLR, BRC, MEM, XX, AJK, and ZB) used a structured data abstraction form and worked independently to abstract key information from full-text journal articles identified by the investigators above. Co-investigators EMT, MLR, BRC, MEM, XX, AJK, and ZB verified data abstractions between them, and disagreements were handled by discussion with the principal investigator PDP. Co-investigators EMT, MLR, BRC, MEM, XX, ZB, AJK, and PDP also searched bibliographies to identify additional relevant research and reviewed those full-text journal articles to determine inclusion or exclusion.

Risk of Bias Assessment

The team's three senior investigators (PDP, MDW, and AF) used the Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) tool to evaluate the risk of bias in retained studies (17). The QUADAS-2 assesses the risk of bias in a study across four domains: patient selection, index test, reference standard, and flow and timing.

Synthesis of Findings

Three senior investigators (PDP, MDW, and AF) categorized the reliability and validity findings of included studies as favorable, unfavorable, or mixed/inconclusive (13, 18).

Criteria for Evaluating Validity

Criterion-related validity (also referred to as predictive validity) was selected as a critical outcome for this systematic review. Criterion-related validity refers to the functional relations between a predictor and criterion events occurring before, during, and after the predictor is applied (19). Predictive validity is often evaluated or assessed by the degree to which the predictor(s) are associated with the criterion (19). A criterion may be operationalized as a 'gold standard' measure or related indicator (i.e., an indirect measure of the criterion). In this study, we do not define a specific cut point for acceptable criterion-related or predictive validity; however, the higher the correlation/association, the more favorable the criterion-related validity (19).

Sensitivity and specificity were selected as important outcomes of interest. For purposes of this review, sensitivity refers to the proportion of shift workers considered to be fatigued by a survey instrument that were also characterized as fatigued by an objective measurement tool or indirect objective measure (e.g., a measure of error or injury). Specificity refers to the proportion of shift workers classified as not fatigued by a fatigue survey and also classified as not fatigued by an objective measure. We did not abstract data from studies to manually calculate sensitivity or specificity.

Criteria for Evaluating Reliability

Internal consistency is a common measure of instrument reliability (20). Internal consistency refers to inter-correlations between items that operationalize an instrument. Cronbach's alpha is a commonly accepted and reported measure of internal consistency (20). Values range from 0 to 1 with values >0.7 considered acceptable internal consistency reliability (21). Low values suggest the instrument may have too few similar or overlapping items (20). Test-retest reliability is another commonly used indicator of instrument reliability (20). While there is no minimum accepted benchmark for test-retest reliability tests, higher values on a 0–1.0 scale are perceived as a positive indicator of test-retest reliability. The procedures for a test-retest assessment involve administering the same instrument to the same target population within a short time period after the initial test and exploring the correlation between instrument scores. There is no standard time period between the initial test and retest, yet two to four weeks is common; some have recommended three or more months between test-retests (20). Some have used

shorter time periods between tests (i.e., several hours to several days) (22, 23). The time selected is dependent on study aims and context for use of the instrument (22, 23).

Quality of Evidence

Three senior investigators PDP, AF, and MDW used the Grading of Recommendations Assessment, Development and Evaluation (GRADE) framework to report on biases in the retained literature, stratified by outcome, and produce a rating of the quality of research as very low, low, moderate, or high (24, 25).

Reporting

The findings are presented in accordance with the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) (26).

RESULTS

The search strategy yielded n = 1,257 unique records (Figure 1). Co-investigators AF and PJC independently screened n = 1,257 titles and abstracts. The interrater agreement for inclusion/exclusion was moderate (Kappa = 0.63). Seventy-seven records were judged to be potentially eligible based on title and abstract alone. Sixty-eight journal articles were identified during the search of bibliographies as potentially relevant and reviewed in full-text format. Thirty-four unique journal articles describing experimental and non-experimental studies were determined to be relevant and key findings were abstracted into tables (Online Supplement Appendix B). One hundred seven journal articles and other literature were excluded with reasons given organized in the Population, Intervention, Comparison, Outcome (PICO) format (Online Supplement Appendix C) (27). The risks of bias assessments associated with the retained research are accessible in Online Supplement Appendix D.

Fatigue

Nineteen studies reported on the reliability and/or validity of ten different fatigue survey instruments (Online Supplement Table 1). The instruments used include: the Brief Fatigue Inventory (BFI); the Swedish Occupational Fatigue Inventory (SOFI); the Fatigue Related Symptoms Questionnaire (F-RSQ); the single-item Crew Status Survey (CSS); the Occupational Fatigue Exhaustion Recovery (OFER) scale; the Chronic Fatigue Scale of the Standard Shiftwork Index (SSI-CFS); the Fatigue Scale (FAS); the Multidimensional Fatigue Inventory (MFI); the Patient Reported Outcomes Measurement Information System (PROMIS) fatigue survey; and the Chalder Fatigue Questionnaire (CFQ).

Seventeen studies reported on reliability and five reported on criterion-related/predictive validity. Only three studies reported on both reliability and validity (Online Supplement Table 1). Most fatigue survey instruments captured in this review solicited perceived fatigue, in general, during the past month and over the previous 7 days or solicited perceptions of chronic fatigue, in general.

Instruments Measuring Fatigue, in General, Over the Past Month and the Past 7 days, and Chronic Fatigue

Criterion-Related/Predictive Validity

Two studies reported criterion-related validity for fatigue survey instruments that referenced feelings of fatigue in general, over the past month or past seven days, or feelings of chronic fatigue (Online Supplement Table 1). One study examined scores on the CFQ against indirect measures of a reference standard (i.e., self-reported injury, medical error, adverse events, or safety compromising behaviors) (3). Findings showed that the odds of self-reported injury, error, or adverse event, and safety-compromising behaviors among the fatigued personnel were increased relative to non-fatigued personnel, (odds ratios of 1.9 (95% CI 1.1, 3.3), 2.2 (95% CI 1.4, 3.3), and 3.6 (95%CI 1.5, 8.3), respectively) (3). The study reported by Barker et al. showed significant ($p < 0.05$) relationship between greater fatigue and worse nursing performance (28).

Reliability

Fourteen studies examined internal consistency (reliability) of survey instruments that assessed fatigue, in general, chronic fatigue, perceived fatigue during the past month, and fatigue over the past 7 days (Online Supplement Table 1). Findings show tests of internal consistency (Cronbach's alpha) values ranged from 0.64 to 0.95 (2, 29). One study reported by Winwood et al. examined test–retest of the OFER (30). Test–retest findings were positive and ranged from 0.61 to 0.69 for the three domains of the OFER tool.

Sensitivity and Specificity

None of the reviewed studies reported on measures of sensitivity or specificity (Online Supplement Table 1).

Instruments Measuring Fatigue in Real-Time, Past 24 hours, or at End of Shift

Criterion-Related/Predictive Validity

Four studies reported criterion-related validity findings for fatigue survey instruments that evaluate

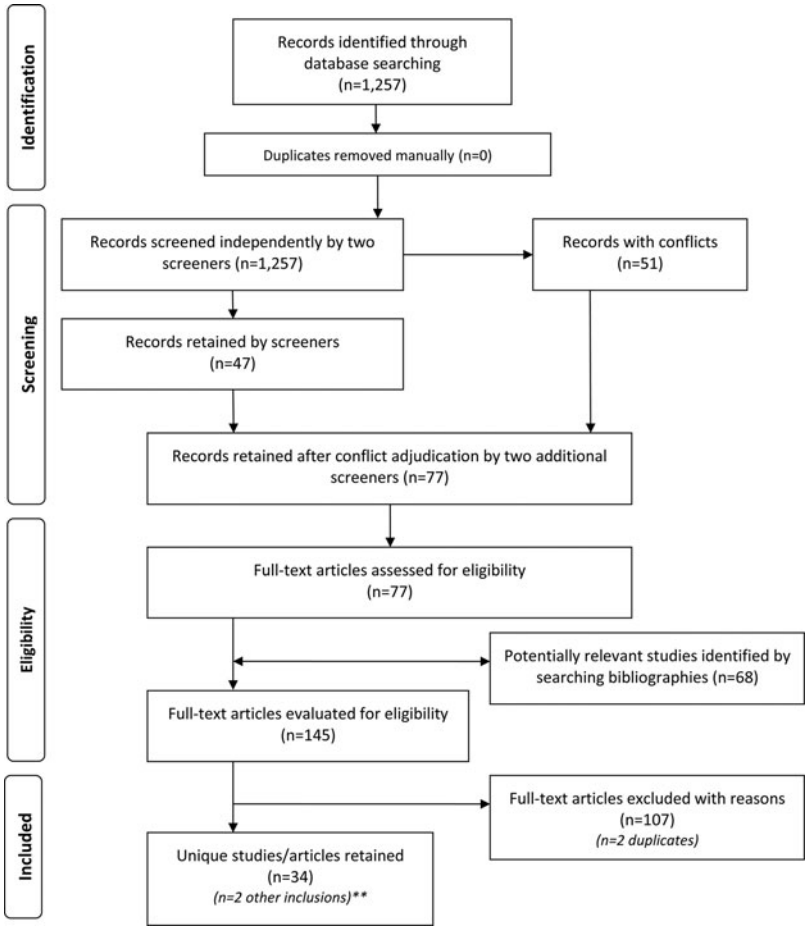

Figure 1. PRISMA Flow Diagram for PICO#1 PROSPERO 2016:CRD42016040097. Notes: ∗∗ Beurskens AJ, Bultmann U, Kant I, Vercoulen JH, Bleijenberg G, Swaen GM. Fatigue among working people: validity of a questionnaire measure. Occup Environ Med. 2000;57(5):353–7. Hossain JL, Reinish LW, Heslegrave RJ, et al. Subjective and objective evaluation of sleep and performance in daytime versus nighttime sleep in extended-hours shift-workers at an underground mine. J Occup Environ Med. 2004;46(3):212–26.

perceived fatigue in real-time, the past 24 hours, or at the end of shift work (Online Supplement Table 1). One study by West et al. used a single item, analog scale adapted from the BFI (31). The respondent indicated feelings of fatigue anchored from "*As bad as it can be*" to "*As good as it can be*." West and colleagues examined the association between fatigue scores on the analog scale and self-reported medical error (an indirect measure of a standard), using multivariable logistic models adjusted for confounders to determine the relationship between fatigue and error. Self-reported medical error was associated with fatigue following adjustment for confounders (See Table 4 in West et al.; 31). Ahsberg et al. examined the association between fatigue measured with the SOFI and reaction time captured at the end of shift work (an indirect measure of a standard) (32). Ahsberg and colleagues detected moderate, yet significant, correlations between four components of the SOFI and reaction time measures at end of shift work (See Table 6 in Ahsberg et al.; 32). Barker and colleagues examined the association between performance scores on the 9-item Nursing Performance Instrument and fatigue as measured by the SOFI, F-RSQ, FAS, and OFER (28). Findings showed consistent significant negative associations, which suggest that as fatigue increases on the 4 fatigue survey instruments, performance decreases (See Table 4 in Barker et al.; 28). Charlton and Bass examined the relationship between scores on the CSS fatigue survey instrument and pass/fail status on a driver simulation test (33). Multivariable modeling revealed that fatigue status was strongly associated with pass/fail status among a sub-set of shift worker participants (i.e., those 37 years of age and younger) (33). The details of the model, including beta coefficients for CSS, were not reported.

Reliability

Three studies examined reliability of a fatigue survey instrument (Online Supplement Table 1). Measures of internal consistency (i.e., Cronbach's alpha) for the SOFI instrument were reported in 2 studies by Ahsberg and colleagues and ranged from 0.68 to 0.93 (32, 34). The reported Cronbach's alpha for SOFI domains reported in the Barker et al. study ranged from 0.76

to 0.89 (28). One study by Grech et al. reported measuring reliability with the "inter-occurrence reliability test" (35). Findings ranged from 0.59 to 0.90.

Sensitivity and Specificity

None of the retained studies reported on measures of sensitivity or specificity for fatigue survey instruments (Online Supplement Table 1).

Sleepiness

Eighteen studies evaluated the reliability and/or validity of four different sleepiness survey instruments (Online Supplement Table 1). Situational sleepiness instruments identified include the Karolinska Sleepiness Scale (KSS), Stanford Sleepiness Scale (SSS), and Sleepiness Symptoms Questionnaire (SSQ). Trait-based sleepiness instruments identified include: (i.e., the Epworth Sleepiness Scale [ESS]). Nine of the retained studies reported on instrument reliability, 10 studies reported criterion-related/predictive validity, and one study reported on both reliability and validity.

Situational Sleepiness Instruments

Criterion-Related/Predictive Validity

Seven studies reported criterion-related validity for sleepiness instruments that measure situational sleepiness (i.e., the KSS, SSS, and SSQ; Online Supplement Table 1). One of 7 studies examined scores on the sleepiness instrument against performance measured on an established standard for sleepiness (e.g., the MSLT) (36). Six studies examined scores on situational sleepiness scales against what may be characterized as objective measures of fatigue (e.g., performance on the psychomotor vigilance task [PVT], driving simulator, postural control, or similar measures; 37–42). The criterion-related/predictive validity findings for 6 of 7 studies were categorized as favorable (Online Supplement Table 1).

Reliability

One study examined test–retest reliability of a situational sleepiness instrument (the Sleepiness Symptoms Questionnaire) (41), and findings were categorized as favorable (Online Supplement Table 1).

Sensitivity and Specificity

None of the retained studies that used a situational sleepiness instrument reported on measures of sensitivity or specificity (Online Supplement Table 1).

Trait Sleepiness Instruments

Criterion-Related/Predictive Validity

Three studies reported criterion-related validity for trait sleepiness instruments that measure situational sleepiness (Online Supplement Table 1). One of three studies examined scores on the trait sleepiness instrument (i.e., the ESS) against performance measured on an established standard for sleepiness (e.g., the Maintenance of Wakefulness Test [MWT]; 43). Two studies examined scores on trait sleepiness instruments against indirect measures of a gold standard (e.g., performance on the PVT, driving simulator, postural control, or similar measures; 31, 44). This research categorized findings for one study as favorable (44), mixed/inconclusive for one study (31), and unfavorable for the last (43) of 3 studies that examined criterion-related/predictive validity for trait sleepiness instruments.

Reliability

Eight studies examined internal consistency reliability of a trait sleepiness instrument (the ESS; Online Supplement Table 1) (45–52). Internal consistency findings for 6 of 8 studies were categorized as favorable. Findings for 2 studies were categorized as unfavorable (50, 51). One study also examined test–retest reliability with the ESS (47). Findings for this analysis were categorized as favorable.

Sensitivity and Specificity

One study reported measuring sensitivity and specificity of the ESS (Online Supplement Table 1) (50). The analysis was interpreted differently and took the findings as a measure of convergent/discriminant validity. We based this interpretation on the observation that investigators did not use an established gold standard measure against which to compare ESS scores. They examined scores on a situational sleepiness instrument against the scores obtained on a trait sleepiness instrument. None of the other studies that used a trait sleepiness instrument evaluated or reported on measures of sensitivity or specificity (Online Supplement Table 1).

Quality of Evidence

Evidence quality was rated very low for all outcomes (Online Supplement Table 2). Most studies were judged to have a serious risk of bias associated with observational study designs such as cross-sectional surveys. Many studies used convenience or non-random sampling. Studies that reported tests of criterion-related validity included indirect measures of a reference standard rather than an actual established standard. Most studies did not sample EMS personnel and thus were downgraded for indirectness. Many studies reported

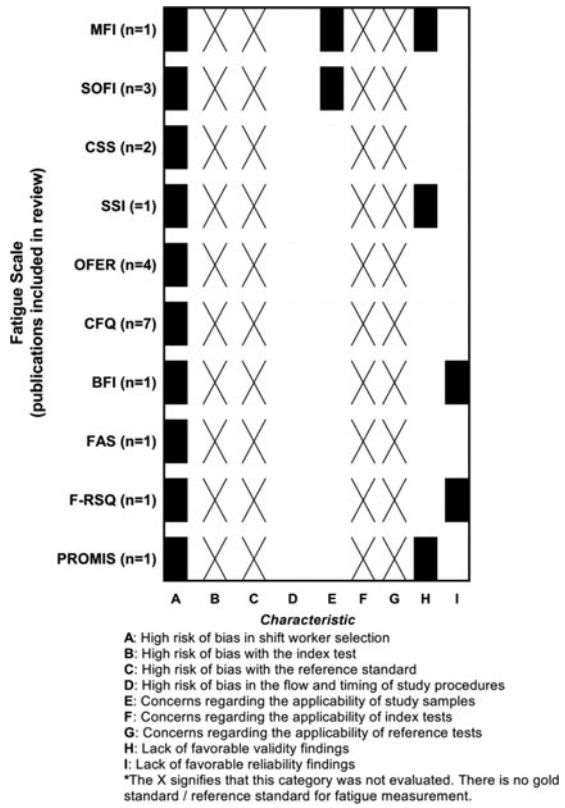

FIGURE 2. Risk of bias, reliability, and validity performance of fatigue survey instruments.

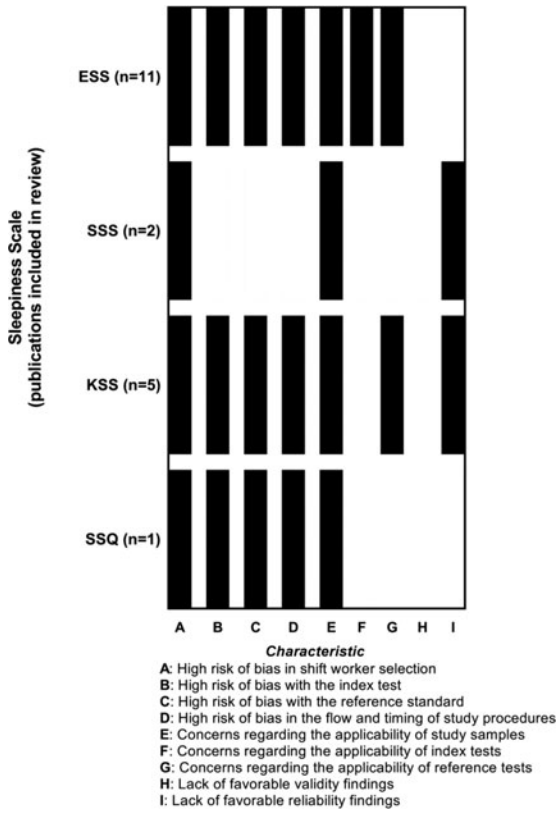

FIGURE 3. Risk of bias, reliability, and validity performance of sleepiness survey instruments.

positive reliability and/or validity findings. It is reasonable to believe that the likelihood of publication might be low if a study reported poor indicators of reliability or validity for fatigue or sleepiness survey instruments. Therefore, the possibility of publication bias was perceived as high.

A summary is presented of instrument utility in graphical format by combining the risk of bias assessment for each survey instrument assessed and the favorability of findings in the research reviewed related to reported findings of reliability and validity (Figures 2 and 3). Nine boxes appear next to each survey instrument and the numbers of studies assessed that report use of a particular instrument appear in parentheses. The graphic includes one box for each of the 7 categories of bias assessed with the QUADAS-2 tool (Refer to Online Supplement Appendix D), one box for favorability (or lack thereof) regarding an instrument's reliability findings, and one box for favorability (or lack thereof) regarding an instrument's validity findings (Refer to Online Supplement Table 1). Black colored boxes signify that the publications/studies assessed, and the application of the survey instrument, were determined to have a high risk of bias and/or the reported reliability and/or the validity findings were not interpreted as favorable. The greater the number of black boxes, the less the confidence in the instrument's utility and use in the EMS population without further testing and research. All interpretations and the messages conveyed in the graphics were guided by application of the GRADE and QUADAS-2 methodologies for assessing bias, and are isolated to the publications in this review. Other sections of this manuscript discuss the potential utility of sleepiness scales that have been widely used in other settings.

DISCUSSION

Summary of Main Results

A limited number of studies reported on the reliability and/or criterion-related validity of fatigue and/or sleepiness survey instruments. None of the reviewed studies tested the sensitivity or specificity of a fatigue or sleepiness survey instrument. Few studies evaluated both reliability and validity. The evidence is limited for the use of these 14 different survey instruments. Further work is needed to determine whether these tools would help EMS administration to measure fatigue and/or sleepiness of EMS personnel.

Inclusion/Exclusion of Prior Research

Many of the reviewed studies used fatigue or sleepiness instruments, yet they did not report on the reliability or validity of survey instruments. These

studies were excluded from this research due to incomplete reporting. In addition, the decision to include or exclude a particular study was based on the study's relevance to the target population, intervention, comparison(s), and outcome(s) of interest. Studies involving healthy volunteers (non-shift workers) were excluded. This study excluded book chapters, theses and dissertations, as well as government reports where the reliability or validity of a survey instrument may have been reported.

Quality of Evidence

Most studies were judged to have serious risk of bias. The potential for publication bias is high given the belief that studies with evidence of poor instrument reliability and/or validity are not widely published.

Agreement and Disagreement with Other Systematic Reviews

Pigeon et al. examined the published literature to draw a distinction between excessive daytime sleepiness and fatigue (53). They searched MEDLINE and the Cochrane Database of Systematic Reviews, examined bibliographies, and contacted authors/investigators. Their search strategy was limited to articles published from 1990 to 2001 and excluded studies that assessed physical fatigue, insomnia, and sleep apnea (with a few exceptions), and work-related fatigue. Shahid et al. reviewed the psychometric properties of the objective and subjective measurements of fatigue and sleepiness (54). They did not document the source databases/repositories or search parameters. Curcio et al. completed a narrative review of survey-based instruments and objective tools for measuring sleepiness (15). Mota and Pimenta did not isolate their review of 18 self-report fatigue instruments to shift workers (55). The systematic review in the present study focused on EMS personnel and related shift worker groups; included a search of multiple databases, as well as literature published from 1980 to September 2016; and was focused on fatigue survey instruments for assessing fatigue in the occupational setting. This systematic review complements previous reviews and enhances the focus on fatigue assessment instruments for EMS personnel and related shift worker groups.

Utility of Fatigue/Sleepiness Survey Instruments Tested in Other Populations

Reliable and valid survey instruments have been developed in other populations. The Fatigue Severity Scale is one such instrument that was initially developed in a clinical population and has since been applied to the general population (56, 57). In addition, the reliability and validity of the CFQ has been demonstrated in occupational settings as well as the general population (5, 58). The Epworth Sleepiness Scale, Karolinska Sleepiness Scale, and Stanford Sleepiness Scale are reliable and valid instruments that are widely used for the assessment of sleepiness (59–61). The ESS was designed to measure trait sleepiness, or habitual sleepiness in recent times, while the KSS and SSS measure situational/state sleepiness, or sleepiness specifically at the time of survey completion. These sleepiness scales are generalized scales that can be readily incorporated as part of an EMS fatigue risk management program. (See Online Supplement Appendix E for copies of instruments examined in this review and published with permission from developers/authors.)

LIMITATIONS

The current study was limited to literature maintained in select databases and searches of reference lists (bibliographies). The judgment of screeners (authors AF and PJC) was examined against decisions of the principal investigator (PDP) with a random sample of n = 50 titles and abstracts pulled from the initial pool of potentially relevant titles and abstracts. The percentage of agreement among AF, PJC, and PDP was 90%. Additional research relevant to this review may exist beyond that identified in this search.

There is no agreed-upon gold standard definition or description of occupational fatigue (1, 6, 7). Frone and Tidwell propose a general definition of work fatigue as: extreme tiredness and reduced functional capacity that is experienced during and at the end of the workday (6). The International Civil Aviation Organization (ICAO) defines fatigue as: A physiological state of mental or physical performance capability resulting from sleep loss or extended wakefulness, circadian phase, or workload (mental/physical activity) that can impair a crew member's alertness and ability to safely operate an aircraft or perform safety-related duties (62). Shen and associates define fatigue as an overwhelming sense of tiredness, lack of energy and feeling of exhaustion, associated with impaired physical and/or cognitive functioning; which needs to be distinguished from symptoms of depression, which include a lack of self-esteem, sadness, and despair or hopelessness (7). Pigeon and colleagues proposed sleepiness be described as drowsiness, sleep propensity, and decreased alertness (53). Pigeon and colleagues described fatigue as weariness, weakness, and depleted energy (53). Relevant studies and literature may have been missed or excluded in this systematic review due to the absence of standardization and inconsistency in defining fatigue.

The study was limited to survey instruments that assessed workplace fatigue in EMS and related shift worker groups. Instruments that measured burnout, sleep quality, and/or sleep disturbance, or intended to screen for insomnia were not included in this review (63). The Maslach Burnout Inventory (MBI), Copenhagen Burnout Inventory (CBI), Standard Shift-

work Index (SSI), Bergen Shift Work Questionnaire (BSWSQ), and similar instruments assess exhaustion, related constructs, and/or label sub-scales as measures of fatigue (6, 49, 64–66). The study herein did not specifically seek out survey instruments that measured burnout, stress, performance, sleep quality, or other constructs. These constructs may be related to fatigue or include a sub-component of fatigue (mental, physical, chronic, or other); yet, the primary focus or intent of the tool was not to measure work-related fatigue. Our study included studies where the focus or aim of the study was the fatigue domain (sub-scale/component) of an instrument that measured multiple constructs (e.g., the SSI and Checklist of Individual Strength Questionnaire [CIS]; 67, 68). If a study used multiple survey instruments, and one or more were relevant to the study (e.g., ESS for sleepiness, the CFQ for fatigue), the reported findings from use of the fatigue and sleepiness survey instrument were retained while ignoring findings from the other instruments (e.g., the BSWSQ; 49). Studies were excluded where the aim was to test the adaptation of a particular tool to a foreign/non-English language (69–71).

Convergent validity, discriminant validity, and content validity were not included as critical or important outcomes of interest. Convergent validity refers to the degree to which an instrument's scores are similar to another instrument measuring the same construct (19, 72). Discriminant validity refers to the scores of an instrument correlating poorly with the scores of another instrument that measures a similar yet different construct (72). Content validity refers to how well an instrument measures the full domain or range of domains being evaluated (73). Construct validity refers to how well the instrument measures the constructs it claims to measure (74). Many studies reported tests of convergent, discriminant, content, and/or construct validity (2, 3, 29–34, 39, 41, 45–47, 49, 50, 67, 68, 75–79). An item-to-construct Pearson correlation value ≥ 0.40, and construct-to-construct Pearson correlation coefficient less than the Cronbach's alpha coefficient value are additional measures of instrument validity (21, 80–83). Several studies reported one or both of these measures or a related measure (e.g., scale/component score-to-total survey score) (2, 34, 75, 76). Two of the retained studies (listed as "other inclusions" in the PRISMA diagram) examined one or more measures of validity other than criterion/predictive related validity and did not report on critical and/or important measures of interest to this study (68, 79).

Conclusions

In this systematic review, limited, though positive, evidence of the reliability and validity of 14 different survey instruments to assess the fatigue and/or sleepiness status of EMS personnel and related shift worker groups were identified. A limited number of generalized tools used commonly in other settings show promise for the assessment of fatigue or sleepiness in EMS. The need for research focused on development and testing of fatigue survey instruments tailored specifically to the EMS operational setting is compelling.

ORCID

Patrick J. Coppler http://orcid.org/0000-0002-0731-7989

References

1. Ream E, Richardson A. Fatigue: a concept analysis. Int J Nurs Stud. 1996;33(5):519–29. doi:10.1016/0020-7489(96)00004-1
2. Patterson PD, Suffoletto BP, Kupas DF, Weaver MD, Hostler D. Sleep quality and fatigue among prehospital providers. Prehosp Emerg Care. 2010;14(2):187–93. doi:10.3109/10903120903524971
3. Patterson PD, Weaver MD, Frank RC, Warner CW, Martin-Gill C, Guyette FX, Fairbanks RJ, Hubble MW, Songer TJ, Callaway CW, et al. Association between poor sleep, fatigue, and safety outcomes in emergency medical services providers. Prehosp Emerg Care. 2012;16(1):86–97. doi:10.3109/10903127.2011.616261
4. Lerman SE, Eskin E, Flower DJ, George EC, Gerson B, Hartenbaum N, Hursh SR, Moore-Ede M. Fatigue risk management in the workplace. J Occup Environ Med. 2012;54(2):231–58. doi:10.1097/JOM.0b013e318247a3b0
5. De Vries J, Michielsen HJ, Van Heck GL. Assessment of fatigue among working people: a comparison of six questionnaires. Occup Environ Med. 2003;60(Suppl 1):i10–i15. doi:10.1136/oem.60.suppl_1.i10
6. Frone MR, Tidwell MC. The meaning and measurement of work fatigue: development and evaluation of the Three-Dimensional Work Fatigue Inventory (3D-WFI). J Occup Health Psychol. 2015;20(3):273–88. doi:10.1037/a0038700
7. Shen J, Barbera J, Shapiro CM. Distinguishing sleepiness and fatigue: focus on definition and measurement. Sleep Med Rev. 2006;10(1):63–6. doi:10.1016/j.smrv.2005.05.004
8. Basner M, Mollicone D, Dinges DF. Validity and sensitivity of a Brief Psychomotor Vigilance Test (PVT-B) to total and partial sleep deprivation. Acta Astronaut. 2011;69(11–12):949–59. doi:10.1016/j.actaastro.2011.07.015
9. Shahid A, Wilkinson K, Marcu S, Shapiro CM. STOP, THAT and One Hundred Other Sleep Scales. New York, NY: Springer Science & Business Media; 2012.
10. Okuyama T, Akechi T, Kugaya A, Okamura H, Shima Y, Maruguchi M, Hosaka T, Uchitomi Y. Development and validation of the cancer fatigue scale: a brief, three-dimensional, self-rating scale for assessment of fatigue in cancer patients. J Pain Symptom Manage. 2000;19(1):5–14. doi:10.1016/S0885-3924(99)00138-4
11. Matza LS, Phillips GA, Revicki DA, Murray L, Malley KG. Development and validation of a patient-report measure of fatigue associated with depression. J Affect Disord. 2011;134(1–3):294–303. doi:10.1016/j.jad.2011.06.028
12. Patterson PD, Higgins JS, Lang ES, Runyon MS, Barger LK, Studnek JR, Moore CG, Robinson K, Gainor D, Infinger A, et al. Evidence-Based Guidelines for Fatigue Risk Management in EMS: Formulating Research Questions and Selecting Outcomes. Prehosp Emerg Care. 2017;21(2):149–56. doi:10.1080/10903127.2016.1241329

13. Patterson PD, Higgins JS, Weiss PM, Lang ES, Martin-Gill C. Systematic review methodology for the Fatigue in Emergency Medical Services project. Prehosp Emerg Care. 2018;22(S1): 9–16.
14. Shadish WR, Cook TD, Campbell DT. Experiments and Generalized Causal Inference. Experimental and Quasi-Experimental Designs for Generalized Causal Inference: Houghton Mifflin Company; 2002.
15. Curcio G, Casagrande M, Bertini M. Sleepiness: evaluating and quantifying methods. Int J Psychophysiol. 2001;41(3):251–63. doi:10.1016/S0167-8760(01)00138-6
16. Guyatt GH, Oxman AD, Kunz R, Atkins D, Brozek J, Vist G, Alderson P, Glasziou P, Falck-Yitter Y, Schunemann HJ. GRADE guidelines: 2. *Framing the question and deciding on important outcomes*. J Clin Epidemiol. 2011;64(4):395–400.
17. Whiting P, Rutjes AW, Reitsma JB, Bossuyt PM, Kleijnen J. The development of QUADAS: A tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. BMC Med Res Methodol. 2003;3:25. doi:10.1186/1471-2288-3-25
18. Bolster L, Rourke L. The effect of restricting residents' duty hours on patient safety, resident well-being, and resident education: an updated systematic review. J Grad Med Educ. 2015;7(3):349–63. doi:10.4300/JGME-D-14-00612.1
19. Nunnally JC, Bernstein IH. Validity. Psychometric theory. 3rd ed. New York, NY: McGraw-Hill, Inc.; 1994:83–113.
20. Nunnally JC, Bernstein IH. The assessment of reliability. Psychometric Theory. 3rd ed. New York, NY: McGraw-Hill, Inc.; 1994:248–92.
21. Hatcher L. A step-by-step approach to using the SAS system for factor analysis and structural equation modeling. Cary, NC: SAS Institute., Inc.; 1994.
22. Dodd K, Hill K, Haas R, Luke C, Millard S. Retest reliability of dynamic balance during standing in older people after surgical treatment of hip fracture. *Physiother Res Int*. 2003;8(2):93–100. doi:10.1002/pri.276
23. Dvir Z, Gal-Eshel N, Shamir B, Prushansky T, Pevzner E, Peretz C. Cervical motion in patients with chronic disorders of the cervical spine: a reproducibility study. Spine. 2006;31(1):E394–E399. doi:10.1097/01.brs.0000219951.79922.df
24. Guyatt GH, Oxman AD, Akl EA, Kunz R, Vist G, Brozek J, Norris S, Falck-Yitter Y, Glasziou P, DeBeer H et al. GRADE guidelines: 1. Introduction-GRADE evidence profiles and summary of findings tables. J Clin Epidemiol. 2011;64(4):383–94. doi:10.1016/j.jclinepi.2010.04.026
25. Guyatt GH, Oxman AD, Vist G, Kunz R, Brozek J, Alonso-Coello P, Montori V, Akl EA, Djulbegovic B, Falck-Ytter Y, et al. GRADE guidelines: 4. Rating the quality of evidence—study limitations (risk of bias). J Clin Epidemiol. 2011;64(4):407–15. doi:10.1016/j.jclinepi.2010.07.017
26. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA-Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. J Clin Epidemiol. 2009;62(10):1006–1010. doi:10.1016/j.jclinepi.2009.06.005
27. Edinger T, Cohen AM. A large-scale analysis of the reasons given for excluding articles that are retrieved by literature search during systematic review. AMIA Annu Symp Proc. 2013:379–87.
28. Barker LM, Nussbaum MA. Fatigue, performance and the work environment: a survey of registered nurses. J Adv Nurs. 2011;67(6):1370–82. doi:10.1111/j.1365-2648.2010.05597.x
29. Friesen LD, Vidyarthi AR, Baron RB, Katz PP. Factors associated with intern fatigue. J Gen Intern Med. 2008;23(12):1981–86. doi:10.1007/s11606-008-0798-3
30. Winwood PC, Lushington K, Winefield AH. Further development and validation of the Occupational Fatigue Exhaustion Recovery (OFER) scale. J Occup Environ Med. 2006;48(4):381–89.
31. West CP, Tan AD, Habermann TM, Sloan JA, Shanafelt TD. Association of resident fatigue and distress with perceived medical errors. JAMA. 2009;302(12):1294–1300. doi: 10.1001/jama.2009.1389
32. Ahsberg E, Kecklund G, Akerstedt T, Gamberale F. Shiftwork and different dimensions of fatigue. Int J Ind Ergon. 2000;26(4):457–65. doi:10.1016/S0169-8141(00)00007-X
33. Charlton SG, Baas PH. Fatigue, work-rest cycles, and psychomotor performance of New Zealand truck drivers. New Zealand Journal of Psychology. 2001;30(1):32–39.
34. Ahsberg E. Dimensions of fatigue in different working populations. Scand J Psychol. 2000;41(3):231–41. doi:10.1111/1467-9450.00192
35. Grech MR, Neal A, Yeo G, Smith S, Humphreys M. An examination of the relationship between workload and fatigue within and across consecutive days of work: is the relationship static or dynamic? J Occup Health Psychol. 2009;14(3):231–42. doi:10.1037/a0014952
36. Chang YS, Wu YH, Su SF, Hsu CY, Chen HL. Factors related with the ability to maintain wakefulness in the daytime after fast and forward rotating shifts. Kaohsiung J Med Sci. 2013;29(11):611–16. doi:10.1016/j.kjms.2013.01.012
37. Balkin TJ, Bliese PD, Belenky G, Sing H, Thorne DR, Thomas M, Redmond DP, Russo M, Wesensten NJ. Comparative utility of instruments for monitoring sleepiness-related performance decrements in the operational environment. J Sleep Res. 2004;13(3):219–27. doi:10.1111/j.1365-2869.2004.00407.x
38. Forsman P, Pyykko I, Toppila E, Haeggstrom E. Feasibility of force platform based roadside drowsiness screening – a pilot study. Accid Anal Prev. 2014;62:186–90. doi: 10.1016/j.aap.2013.09.015
39. Geiger-Brown J, Wieroney M, Blair L, Zhu S, Warren J, Scharf SM, Hinds PS. Measuring subjective sleepiness at work in hospital nurses: Validation of a modified delivery format of the Karolinska Sleepiness Scale. Sleep Breath. 2014;18(4):731–39. doi:10.1007/s11325-013-0935-z
40. Karchani M, Mazloumi A, NaslSaraji G, Akbarzadeh A, Niknezhad A, Ebrahimi MH, Raei M, Khandan M. Association of subjective and interpretive drowsiness with facial dynamic changes in simulator driving. J Res Health Sci. 2015;15(4): 250–55.
41. Howard ME, Jackson ML, Berlowitz D, O'Donoghue F, Swann P, Westlake J, Wilkinson V, Pierce RJ. Specific sleepiness symptoms are indicators of performance impairment during sleep deprivation. Accid Anal Prev. 2014;62:1–8. doi:10.1016/j.aap.2013.09.003
42. Ftouni S, Sletten TL, Howard M, Anderson C, Lenne MG, Lockley SW, Rajaratnam SM. Objective and subjective measures of sleepiness, and their associations with on-road driving events in shift workers. J Sleep Res. 2013;22(1):58–69. doi:10.1111/j.1365-2869.2012.01038.x
43. Sadeghniiat-Haghighi K, Moller HJ, Saraei M, Aminian O, Khajeh-Mehrizi A. The Epworth Sleepiness Scale for screening of the drowsy driving: comparison with the maintenance of wakefulness test in an Iranian sample of commercial drivers. Acta Med Iran. 2014;52(2):125–29.
44. Sunwoo BY, Jackson N, Maislin G, Gurubhagavatula I, George CF, Pack AI. Reliability of a single objective measure in assessing sleepiness. Sleep. 2012;35(1):149–58. doi:10.5665/sleep.1606
45. Sofianopoulos S, Williams B, Archer F, Thompson B. The exploration of physical fatigue, sleep and depression in paramedics: a pilot study. eJournal of Emergency Primary Health Care. 2011;9(1):1–33.
46. Waage S, Pallesen S, Moen BE, Mageroy N, Flo E, Di Milia L, Bjorvatn B. Predictors of shift work disorder among nurses: a longitudinal study. Sleep Med. 2014;15(12):1449–55. doi:10.1016/j.sleep.2014.07.014

47. Johns MW. Reliability and factor analysis of the Epworth Sleepiness Scale. Sleep. 1992;15(4):376–81. doi:10.1093/sleep/15.4.376
48. Saksvik-Lehouillier I, Bjorvatn B, Hetland H, Sandal GM, Moen BE, Mageroy N, Akerstedt T, Pallesen S. Individual, situational and lifestyle factors related to shift work tolerance among nurses who are new to and experienced in night work. J Adv Nurs. 2013;69(5):1136–46. doi:10.1111/j.1365-2648.2012.06105.x
49. Flo E, Bjorvatn B, Folkard S, Moen BE, Gronli J, Nordhus IH, Pallesen S. A reliability and validity study of the Bergen Shift Work Sleep Questionnaire in nurses working three-shift rotations. Chronobiol Int. 2012;29(7):937–46. doi:10.3109/07420528.2012.699120
50. Kato C, Shimada J, Hayashi K. Sleepiness during shift work in Japanese nurses: a comparison study using JESS, SSS, and actigraphy. Sleep and Biological Rhythms. 2012;10:109–17. doi:10.1111/j.1479-8425.2011.00528.x
51. Takahashi M, Nakata A, Haratani T, Otsuka Y, Kaida K, Fukasawa K. Psychosocial work characteristics predicting daytime sleepiness in day and shift workers. Chronobiol Int. 2006;23(6):1409–22. doi:10.1080/07420520601100963
52. Scott LD, Hofmeister N, Rogness N, Rogers AE. An interventional approach for patient and nurse safety: a fatigue countermeasures feasibility study. Nurs Res. 2010;59(4):250–58. doi:10.1097/NNR.0b013e3181de9116
53. Pigeon WR, Sateia MJ, Ferguson RJ. Distinguishing between excessive daytime sleepiness and fatigue: toward improved detection and treatment. J Psychosom Res. 2003;54(1):61–69. doi:10.1016/S0022-3999(02)00542-1
54. Shahid A, Shen J, Shapiro CM. Measurements of sleepiness and fatigue. J Psychosom Res. 2010;69(1):81–9. doi:10.1016/j.jpsychores.2010.04.001
55. Mota DD, Pimenta CA. Self-report instruments for fatigue assessment: a systematic review. Res Theory Nurs Pract. 2006;20(1):49–78. doi:10.1891/rtnp.20.1.49
56. Krupp LB, LaRocca NG, Muir-Nash J, Steinberg AD. The fatigue severity scale. Application to patients with multiple sclerosis and systemic lupus erythematosus. Arch Neurol. 1989;46(10):1121–23. doi:10.1001/archneur.1989.00520460115022
57. Beck AT, Brown G, Steer RA. Prediction of eventual suicide in psychiatric inpatients by clinical ratings of hopelessness. J Consult Clin Psychol. 1989;57(2):309–10. doi:10.1037/0022-006X.57.2.309
58. Loge JH, Ekeberg O, Kaasa S. Fatigue in the general Norwegian population: normative data and associations. J Psychosom Res. 1998;45(1):53–65. doi:10.1016/S0022-3999(97)00291-2
59. Johns MW. A new method for measuring daytime sleepiness: The Epworth sleepiness scale. Sleep. 1991;14(6):540–45. doi:10.1093/sleep/14.6.540
60. Akerstedt T, Gillberg M. Subjective and objective sleepiness in the active individual. Int J Neurosci. 1990;52(1–2):29–37. doi:10.3109/00207459008994241
61. Hoddes E, Zarcone V, Smythe H, Phillips R, Dement WC. Quantification of sleepiness: a new approach. Psychophysiology. 1973;10(4):431–36. doi:10.1111/j.1469-8986.1973.tb00801.x
62. Caldwell JA, Caldwell JL. An overview of fatigue. Fatigue in aviation: A guide to staying awake at the stick. 2nd ed. New York, NY: Routledge Taylor and Francis Group; 2016:15.
63. Jeong HS, Jeon Y, Ma J, Choi Y, Ban S, Lee S, Lee B, Im JJ, Yoon S, Kim JE, et al. Validation of the Athens Insomnia Scale for screening insomnia in South Korean firefighters and rescue workers. Qual Life Res. 2015;24(10):2391–95. doi:10.1007/s11136-015-0986-7
64. Lee KA, Gay CL, Alsten CR. Home-based behavioral sleep training for shift workers: a pilot study. Behav Sleep Med. 2014;12(6):455–68. doi:10.1080/15402002.2013.825840
65. Halbesleben JRB. The influence of shift work on emotional exhaustion in firefighters: the role of work–family conflict and social support. *Int J Workplace Health Management*. 2009;2(2):115–30. doi:10.1108/17538350910970200
66. Barton J, Spelten E, Totterdell P, Smith L, Folkard S, Costa G. The Standard Shiftwork Index: a battery of questionnaires for assessing shiftwork-related problems. Work Stress. 1995;9(1):4–30. doi:10.1080/02678379508251582
67. Ruggiero JS. Correlates of fatigue in critical care nurses. Res Nurs Health. 2003;26(6):434–44. doi:10.1002/nur.10106
68. Beurskens AJ, Bultmann U, Kant I, Vercoulen JH, Bleijenberg G, Swaen GM. Fatigue among working people: validity of a questionnaire measure. Occup Environ Med. 2000;57(5):353–57. doi:10.1136/oem.57.5.353
69. Laranjeira CA. Translation and adaptation of the fatigue severity scale for use in Portugal. Appl Nurs Res. 2012;25(3):212–17. doi:10.1016/j.apnr.2010.11.001
70. Guttierrez JL, Jimenez BM, Hernandez EG, Lopez AL. Spanish version of the Swedish Occupational Fatigue Inventory (SOFI): Factorial replication, reliability and validity. Int J Ind Ergon. 2005;35(8):737–46. doi:10.1016/j.ergon.2005.02.007
71. Chica-Urzola HL, Escobar-Cordoba F, Eslava-Schmalbach J. Validating the Epworth Sleepiness Scale. Rev Salud Publica (Bogota). 2007;9(4):558–67.
72. Campbell DT, Fiske DW. Convergent and discriminant validation by the multitrait-multimethod matrix. Psychol Bull. 1959;56(2):81–105. doi:10.1037/h0046016
73. Rubio DM, Berg-Weger M, Tebb SS, Lee ES, Rauch S. Objectifying content validity: conducting a content validity study in social work research. Soc Work Res. 2003;27(2):94–104. doi:10.1093/swr/27.2.94
74. Smith GT. On construct validity: issues of method and measurement. Psychol Assess. 2005;17(4):396–408. doi:10.1037/1040-3590.17.4.396
75. Patterson PD, Buysse DJ, Weaver MD, Suffoletto BP, McManigle KL, Callaway CW, Yealy DM. Emergency healthcare worker sleep, fatigue, and alertness behavior survey (SFAB): development and content validation of a survey tool. Accid Anal Prev. 2014;73:399–411. doi:10.1016/j.aap.2014.09.028
76. Patterson PD, Buysse DJ, Weaver MD, Callaway CW, Yealy DM. Recovery between work shifts among emergency medical services clinicians. Prehosp Emerg Care. 2015;19(3):365–75. doi:10.3109/10903127.2014.995847
77. da Silva FJ, Felli VE, Martinez MC, Mininel VA, Ratier AP. Association between work ability and fatigue in Brazilian nursing workers. Work. 2015;53(1):225–32. doi:10.3233/WOR-152241
78. Smets EM, Garssen B, Bonke B, De Haes JC. The Multidimensional Fatigue Inventory (MFI) psychometric qualities of an instrument to assess fatigue. J Psychosom Res. 1995;39(3):315–25. doi:10.1016/0022-3999(94)00125-O
79. Hossain JL, Reinish LW, Heslegrave RJ, Hall GW, Kayumov L, Chung SA, Bhuiya P, Jaovanovic D, Huterer N, Volkov J, et al. Subjective and objective evaluation of sleep and performance in daytime versus nighttime sleep in extended-hours shift-workers at an underground mine. J Occup Environ Med. 2004;46(3):212–26. doi:10.1097/01.jom.0000117421.95392.31
80. Lance CE, Butts MM, Michels LC. The sources of four commonly reported cutoff criteria: what did they really say? Organizational Res Methods. 2006;9(2):202–20. doi:10.1177/1094428105284919
81. Anderson JC, Gerbing DW. Structural equation modeling in practice: a review and recommended two-step approach. Psychol Bull. 1988;103(3):411–23. doi:10.1037/0033-2909.103.3.411
82. Marsh HW, Balla JR, McDonald RP. Goodness-of-fit indexes in confirmatory factor analysis: The effect of sample size. Psychol Bull. 1988;103(3):391–410. doi:10.1037/0033-2909.103.3.391
83. Ware JEJ, Gandek B. Methods for testing data quality, scaling assumptions, and reliability: the IQOLA Project approach. International Quality of Life Assessment. J Clin Epidemiol. 1998;51(11):945–52. doi:10.1016/S0895-4356(98)00085-7