---

**Article:**
Houdmont, J., Randall, R. orcid.org/0000-0002-6991-9037, Kinman, G. et al. (3 more authors) (2021) Can a single-item measure of job stressfulness identify common mental disorder? International Journal of Stress Management, 28 (4). pp. 305-313. ISSN 1072-5245

https://doi.org/10.1037/str0000231

---

Accepted 26th May 2021 for publication in the International Journal of Stress Management.

# Can a Single-Item Measure of Job Stressfulness Identify Common Mental Disorder?

Jonathan Houdmont[1], Raymond Randall[2], Gail Kinman[3], Jim Colwell[4], Robert Kerr[5], and Ken Addley[5]

[1]School of Medicine, University of Nottingham

[2]Sheffield University Management School, University of Sheffield, and Loughborough University Business School, Loughborough University

[3]Department of Organizational Psychology, Birkbeck, University of London

[4]Devon and Cornwall Police

[5]Department of Management, Leadership, and Marketing, Ulster University

## Author Note

Jonathan Houdmont https://orcid.org/0000-0001-5174-2619

Raymond Randall https://orcid.org/0000-0002-6991-9037

Gail Kinman https://orcid.org/0000-0002-0130-1708

Jim Colwell https://orcid.org/0000-0001-5766-4674

Robert Kerr https://orcid.org/0000-0002-1523-979X

Ken Addley https://orcid.org/0000-0002-8177-503X

**Abstract**

There is a need for brief and non-intrusive measures to identify common mental disorder (CMD) in worker populations. The primary aim of this study was to determine whether workers reporting CMD symptoms indicative of minor psychiatric morbidity could be reliably identified by a single-item job stressfulness measure (SIJSM). A secondary aim was to determine the number of response categories required to maximize the sensitivity and specificity of the SIJSM. Data from seven occupational groups were analysed ($N$ = 20,658). We measured CMD using the 12-item General Health Questionnaire (GHQ-12) and job stressfulness with a single item involving five response options. We applied tests of discriminatory power to assess whether a report of high job stressfulness (SIJSM score ≥4, *very stressful* or *extremely stressful*) correctly classified CMD cases (GHQ-12 score ≥4) and non-cases. Both sensitivity and specificity of the SIJSM were acceptable (≥70%) in samples where at least 50% of respondents reported high job stressfulness (prison officers, public protection unit police officers dealing with domestic violence and child abuse). Discriminatory power was optimal and almost identical at the ≥4 cut-off on a 5-point scale and ≥6 on a 9-point scale. In occupations with elevated prevalence of high job stressfulness the SIJSM appears to demonstrate acceptable sensitivity and specificity, providing for efficient and non-intrusive identification of likely minor psychiatric morbidity. The measure could be used with such groups to identify workers that would benefit from in-depth psychosocial risk assessment and targeted intervention.

*Keywords*: concurrent validity, GHQ-12, minor psychiatric morbidity, sensitivity and specificity, stress

## Can a Single-Item Measure of Job Stressfulness Identify Common Mental Disorder?

Single-item measures of the antecedents, outcomes, and overall experience of job stress are widely used in occupational health psychology research and practice. These non-intrusive measures can be administered at regular intervals with minimal disruption. A single-item measure of job stress that reliably identifies workers experiencing psychological health problems could be used to identify groups requiring in-depth psychosocial risk assessment and targeted intervention. There remains, however, a paucity of evidence on the concurrent validity of such measures in relation to established measures of common mental disorder (CMD). The objective of this study is to test the extent to which responses on a single-item measure of global job stressfulness reliably discriminate between workers reporting a case or non-case of CMD indicative of minor psychiatric morbidity.

### Applications of Single-item Measures of Stress

Single-item measures of stress consistently show concurrent relations with established self-report measures of mental wellbeing in worker populations. Tested measures have focused on working conditions that may give rise to the experience of stress (Gilbert & Kelloway, 2014), the emotional experience of stress (Arapovic-Johansson et al., 2017; Elo et al., 2003), and global job stressfulness (Houdmont et al., 2021; Scottish Government, 2018). Single-item measures might therefore prove useful in organizational psychosocial risk management activities. Such measures could be used to provide an indirect assessment of mental wellbeing that is brief, involves minimal disruption to work activities and demand on workers, and generate data that are straightforward to analyse and interpret. Moreover, such measures may be administered at regular intervals, especially via mobile and wearable technology, to enable the continual monitoring of stress, inform the allocation of organizational resources to coincide with peak periods, and identify workgroups at high risk of developing stress-related problems. Their brevity means they could also be used repeatedly

to enable and enhance the evaluation of stress-reduction interventions. Potential benefits and applications of such measures were illustrated in a Swedish study in which primary healthcare staff received a weekly text message containing a single-item measure centred on the experience of stress: “Stress means a state in which a person feels tense, restless, nervous or anxious or is unable to sleep at night because his/her mind is troubled all the time. Do you feel this kind of stress these days?” - with responses given on a 5-point scale of *not at all* (1) to *very much* (5) (Arapovic-Johansson et al., 2017). In tests of convergent validity significant correlations in the expected direction were observed with established multi-item measures of exhaustion ($r = .58$, $p < .01$) and depression ($r = .45$, $p < .01$). Moreover, the single-item measure predicted sickness absence, exhaustion, and depression at 12-month follow-up. Weekly administration of the measure also revealed variability over time in reports of stress, with higher levels in the lead up to Christmas (Arapovic-Johansson et al., 2020). These findings point to the efficacy of a single-item measure for identifying early signs of symptoms that could develop into CMD with implications for work attendance and temporal variations in stress-related issues that could be used to make decisions about the deployment of organizational resources.

The extant literature highlights scope for the application of single-item measures of stress in organizational psychosocial risk management activities. However, the scientific literature contains numerous measures that address contrasting elements of the stress process. This has resulted in knowledge gaps regarding question stem wording and response options needed in a single-item measure of stress required to optimize associations with CMD in worker populations. It is possible that concurrent linkages between a single-item measure of stress and health outcomes might be optimized where a single-item is both explicitly focused on work and global in the sense of permitting respondents to consider personally salient factors in their determination of a response. In this paper we focus on one such single-item

job stressfulness measure (SIJSM). This focus stems from the evidence concerning the superiority of role-and context-specific job stressors over generic ones in the prediction of stress-related outcomes (Brough & Biggs, 2015).

Responses to a SIJSM consistently demonstrate an inverse correlation with mental wellbeing (e.g., Houdmont et al., 2021; Scottish Government, 2018). Yet there remains a paucity of knowledge on the extent to which such a measure is capable of accurately identifying CMD cases indicative of minor psychiatric disorder. This is an important knowledge gap since the identification of CMD cases gives rise to a powerful imperative for further in-depth assessment and intervention activities. High discriminatory power in a SIJSM would further give rise to several research applications. Brevity in assessment of CMD may be attractive to researchers keen to reduce survey fatigue, participant frustration that can be associated with lengthy surveys, and interruption to work activities. A SIJSM with high discriminatory power in relation to CMD would allow for repeated administration with good response rates and low attrition, thereby facilitating longitudinal research and the robust evaluation of the outcomes of stress-reduction interventions (Johnson et al., 2018). A SIJSM could also be used in lengthy workforce survey instruments that measure in detail an array of constructs and where an imperative exists to assess CMD while minimizing the time required for survey completion. Finally, a measure with high discriminatory power could support the efficient pre-screening of study participants who could then be assessed for CMD problems of clinical severity. Therefore, the first aim of this study is to assess the extent to which a SIJSM can identify cases and non-cases of CMD indicative of minor psychiatric morbidity.

**Single-item Measure of Job Stress**

The SIJSM that we examine in this study was developed for the Bristol Stress and Health at Work (SHAW) study (Smith, 2000, 2001; Smith et al., 2000) commissioned by the

United Kingdom's (UK) Health and Safety Executive[1] to determine the scale of perceived stress at work in a random population sample. It involves the stem question 'In general, how do you find your job?' with five labelled response alternatives: *not at all stressful* (1), *mildly stressful* (2), *moderately stressful* (3), *very stressful* (4), and *extremely stressful* (5). High job stressfulness is defined as a response of *very stressful* or *extremely stressful* on the basis that: "First, we consider that no organisation would want their employees to be very stressed. Second, it is common practice to define your 'high' group as the upper quartile, and our estimate of the prevalence of perceived occupational stress falls close to this figure" (Smith et al., 2000, p. 212).

This measure is widely used in large-scale workforce surveys particularly, though not exclusively, in the UK (e.g., Elliott-Davies, 2018, 2021; Health and Safety Executive, 2012; Hodgson et al., 2006; Scottish Government, 2018) and academic research (e.g., Allisey et al., 2014; Blake et al., 2020; Eng et al., 2011; Houdmont et al., 2021; Houdmont et al., 2012; Phillips et al., 2008). Qualitative examination of the measure's construct validity has identified working conditions - primarily job demands - as the primary frame of reference considered by workers when forming a response (Houdmont et al., 2019).

The SHAW measure closely resembles that featured in large-scale influential workforce survey series in North America and Europe. For instance, since 2001 the nationally representative annual Canadian Community Health Survey has included a single-item measure of work-related stress that has contributed to the North American scientific literature (Lachance et al., 2020; Szeto & Dobson, 2013). The SHAW and Canadian measures involve identically labelled endpoints on a 5-point rating scale (*not at all stressful* and *extremely stressful*), while the remaining points have slightly different labels (e.g., *mildly stressful* vs. *not very stressful*). Similarities are also evident with the measure used in the

[1] The UK government agency responsible for regulation of work-related health and safety.

European Working Conditions Survey (EWCS) that since 1991 has been used to collect data from a representative sample of the European workforce approximately every five years. The sixth EWCS (EuroFound, 2016) involved 43,850 workers across 35 countries who indicated their degree of agreement with the statement 'You experience stress in your work' on a 5-point scale of: *always* (1), *most of the time* (2), *sometimes* (3), *rarely* (4), and *never* (5).

The applications of a single-item measure provide some indication of the widespread need for brief measures of job stress. Yet despite growth in the popularity of single-item measures, some wariness remains. Unease arises primarily from two concerns: that single-item measures may not satisfactorily represent the full content domain of conceptually complex constructs and may be unreliable because they do not allow for the assessment of internal consistency (Fisher et al., 2016). Some rebuttal to the first of these points can be found in research addressing constructs such as job satisfaction where a global single-item measure can have better content validity than a multi-item measure assessing several facets of the construct (Scarpello & Campbell, 1983), In the study of job stress, a global single-item measure arguably provides better coverage of the domain than a multi-item measure since it permits respondents to consider personally salient factors rather than a predetermined list of presumed relevant factors provided by the researcher (Houdmont et al., 2019). On the second point, though single-item measures present a challenge for the assessment of response consistency, the absence of an inter-item coefficient may become less of a concern where a single-item measure demonstrates predictive or concurrent validity equal to that achieved by a multi-item measure with high internal reliability (Bergkvist & Rossiter, 2007). Moreover, a single-item measure represents a pragmatic alternative to a multi-item measure where practical constraints would mean that the construct of interest would otherwise go unmeasured (Fisher et al., 2016). Indeed, a single-item measure of job stress could be used in a regular and ongoing cycle of non-intrusive psychosocial risk assessment that might

otherwise be unfeasible owing to concerns about disruption to work activities arising from use of multi-item measures. Brevity of assessment also makes single-item measures attractive for use in research involving experience sampling methods and ecological momentary assessment characterised by self-reports of current or very recent behavior, cognitions, or emotion in real-world settings on multiple occasions over time (Trull & Ebner-Priemer, 2009).

**Optimising the Number of Response Categories in Rating Scales**

The Likert-type rating scale in Smith et al.'s (2000) SIJSM has five response alternatives, with high job stressfulness defined as a response at or above the fourth point. Research on the optimisation of the number of response alternatives, largely located in the consumer marketing literature (e.g., Dawes, 2008), suggests that seven to 10 response alternatives may produce the highest test-retest reliability coefficients and scales with nine response alternatives the highest criterion-related validity (Preston & Colman, 2000). Similarly, studies in educational psychology have shown that test-retest reliability may be optimised with 6 to 9 response categories (Weng, 2004). In view of these findings, we examined whether an extended 9-point rating scale that permits respondents to make finer-grained distinctions in their level of perceived job stressfulness might offer superior discriminatory power in relation to the identification of CMD. Thus, the second aim of the current investigation is to examine whether a 9-point rating scale offers superior discriminatory power when compared to Smith et al.'s (2000) original 5-point rating scale for the identification of CMD.

**Purpose of the Present Study**

We sought to contribute to knowledge by examining the properties of a widely used single-item measure that is explicitly and solely focused on the work domain and quantifies perceptions of overall job stressfulness. We tested whether this measure could be used to

identify concurrent self-reported CMD indicative of minor psychiatric morbidity in a range of occupational groups. If responses to the measure identify CMD cases and non-cases with acceptable sensitivity and specificity, this would lend support to its use in organizational psychosocial risk management activities and research where an imperative exists for efficient, repeatable, and non-intrusive assessment of the prevalence of clinical-level CMD in worker populations.

## Method

### Participants

This study used cross-sectional employee survey data contributed by workers drawn from seven occupational groups. This included office-based civil servants employed by the Northern Ireland Civil Service ($n$ = 16,495) and UK prison officers ($n$ = 1,837) responsible for the supervision, management, and control of prisoners in a lawful, safe, and secure manner. We also included UK police officers employed in several roles. These included public protection unit officers ($n$ = 452) who routinely deal with complex and high stakes issues including child protection, neglect and abuse investigation, management of dangerous offenders, domestic abuse, rape and sexual assault, and protection of vulnerable adults; first response officers ($n$ = 497) who are the first line of response to an incident where police attendance is required and whose job it is to assess the situation; crime investigation officers ($n$ = 215), rural policing team officers ($n$ = 260), and police custody officers ($n$ = 902) who manage a custody suite with responsibility for the care and welfare of detained persons and the decision to authorize or refuse the detention of persons presented before them. All data were drawn from workforce health assessment studies commissioned by the host organizations or employee representative bodies, details of which are reported elsewhere (Addley et al., 2006; Kinman & Clements, 2021; Houdmont, 2014; Houdmont et al., 2021;

Houdmont & Randall, 2016). The socio-demographic composition of each sample is shown in Table 1.

**Measures**

*Job stressfulness*: To measure perceived global job stressfulness, we used the SIJSM developed for the Bristol Stress and Health at Work (SHAW) study (Smith, 2000, 2001; Smith et al., 2000) described in the introduction. When testing the concurrent discriminant validity of the measure, we dichotomised scores into high job stressfulness (*very stressful, extremely stressful*) and low job stressfulness (*not at all stressful, mildly stressful, moderately stressful*) categories (Smith et al., 2000).

To test the optimum number of response categories required to maximize sensitivity and specificity we developed a variant of the SIJSM involving the same stem and an expanded 9-point rating scale. This was completed by participants in the rural policing teams police officer sample ($n$ = 260). The 9-point rating scale was end-anchored with *not at all stressful* (1) on the left and *extremely stressful* (9) on the right and was presented to participants as a series of numbers spaced out. The third, fifth, and seventh points were labelled *mildly stressful*, *moderately stressful*, and *very stressful* respectively, using wording identical to that on the five-point SIJSM. The selection of a 9-point scale was informed by evidence suggesting that eight or nine response categories may maximize reliability (Preston & Colman, 2000). Moreover, it allowed labels to be evenly spread out across the rating scale at alternate points while also end-anchored.

*Common mental disorder*: To assess general (non-psychotic) CMD we used the 12-item General Health Questionnaire (GHQ-12) (Goldberg & Williams, 1988). The GHQ-12 offers a context-free assessment of symptoms of anxiety, depression, social dysfunction, and loss of confidence. Each item requires respondents to consider whether they have recently experienced a behavior or symptom more or less than usual. Negatively framed items (e.g.,

felt constantly under strain) are scored on a 4-point scale of *not at all* (0), *no more than usual* (0), *rather more than usual* (1), and *much more than usual* (1), while response categories for positively framed items (e.g., been able to face up to problems) are *more so than usual* (0), *same as usual* (0), *less than usual* (1), *much less than usual* (1). We used the GHQ (binary/standard) scoring method (0–0–1–1) to identify individuals reporting sufficient psychological distress to be a probable case of minor psychiatric morbidity, with responses summed to a global score ranging from 0 to 12 and dichotomized into non-distressed (GHQ score 0–3) and distressed (GHQ score 4–12). The 3/4 threshold has been identified as providing the optimum balance between sensitivity and specificity for the identification of likely cases of minor psychiatric morbidity in UK occupational (Hardy et al., 1999) and primary care (Goldberg et al., 1997, 1998) samples, and is widely used in UK occupational health research (Goodwin et al., 2013).

**Analytical Approach**

To assess the performance of the SIJSM in relation to CMD, we first applied a Pearson's ($r$) correlation between the continuous values to examine strength and direction of association between the two variables. We applied several tests of discrimination between the measures (Trevathan, 2017). The odds of CMD caseness associated with a report of high job stressfulness, relative to a report of low job stressfulness, were assessed using odds ratios with 95% confidence intervals. Sensitivity was examined by testing the proportion of cases of CMD correctly identified by the SIJSM (true positive rate). Specificity was examined by testing the proportion of CMD non-cases correctly identified by the SIJSM (true negative rate). Positive and negative predictive values (PPV and NPV) were calculated to estimate the odds that participants reporting high job stressfulness also report CMD caseness (PPV) and the probability that those with low job stressfulness also do not report CMD caseness (NPV). Positive and negative likelihood ratios (LR+ and LR-) were calculated to evaluate the

probability that the SIJSM produced a true positive and true negative result respectively. We also conducted a receiver operator curve (ROC) analysis to calculate the area under the curve (AUC). This assessed both the sensitivity and specificity of the ≥4 cut-off on the SIJSM to correctly identify cases of CMD. ROC curves plot a measure's true positive rate (sensitivity) and false positive rate (1-specificity). A test with perfect sensitivity and specificity is indicated by an AUC of 1.0. Our interpretation of AUC values follows the recommendations of Hosmer and Lemeshow (2000) whereby .70 to .79 indicates acceptable discrimination, .80 to .89 excellent discrimination, and ≥.90 outstanding discrimination. These thresholds are widely used in research concerning the discriminant validity of measures of mental wellbeing (e.g., Lundin et al., 2016). Analyses were conducted separately for each occupational group because of the marked variation in high job stressfulness and CMD caseness across occupations (see Table 1). Following Smith et al.'s (2000) original recommendation, high job stressfulness is typically defined as a response of ≥4 (the job being very stressful or extremely stressful). To explore whether discriminatory power of the measure might be improved at the ≥3 (moderately stressful) and 5 (extremely stressful) cut-offs we calculated sensitivity, specificity, and the AUC. We defined the optimal cut-off as that which produced the highest rate of correct classifications and largest AUC. To identify the optimum cut-off score on the 5- and 9-point rating scale we calculated discriminatory statistics for each possible cut-off. All analyses were conducted in IBM SPSS version 24.

## Results

### Personal Characteristics and Prevalence Statistics

Personal characteristics and prevalence statistics are reported by occupational role in Table 1. Participants were mostly male in five occupational groups (68-87%), whereas most civil servants (52%) and public protection unit police officers (59%) were female. Mean age ranged from 39 years (*SD* = 10.79) (civil servants) to 48 (*SD* = 9.44) (prison officers). Civil

servants reported the lowest prevalence of high job stressfulness and CMD caseness (18% and 27% respectively), while prison officers reported the highest prevalence of both states (53% and 59%). Public protection unit police officers also reported a high prevalence of both states (50% and 56%), while the remaining samples consisting of police officers in four different roles (rural, response, crime investigation, custody) reported prevalence rates within a narrower range: high job stressfulness, 30-39%; CMD caseness, 44-56%.

[insert Table 1 about here]

**Identification of CMD**

Association and discriminatory statistics are presented in Table 1. Pearson's correlation coefficients between job stressfulness and CMD were moderate to strong, providing some preliminary evidence of concurrent validity (Cohen, 1988). Odds ratios indicated that workers reporting high job stressfulness were five times (custody police officers) to 10 times (police response officers) more likely to report CMD caseness than those reporting low job stressfulness. Sensitivity and specificity were both acceptable (≥70%) in the prison officer and public protection unit police officer samples. Notably, in these samples ≥50% reported high job stressfulness while CMD caseness prevalence was not markedly higher than observed among the other occupational samples. Sensitivity was 72% (prison officers) and 71% (public protection unit police officers), indicating that the SIJSM identified seven out of 10 individuals with a case of CMD (true positives) while three out of 10 of those with CMD caseness went undetected (false negatives). Specificity was 74% and 78% respectively in these samples, indicating that three quarters of those without a case of CMD were correctly identified (true negatives) and one quarter were incorrectly identified as a CMD case (false positives). In the remaining five samples, sensitivity was lower, ranging from 42% to 61%, indicating a comparatively high false negative rate; contrastingly, specificity was high across these samples (81-91%), indicating a consistently low false

positive rate. The area under the ROC curve was ≥.70 in six of the samples, indicating reasonable discrimination, and just below this threshold in the police custody officer sample (.67).

We tested whether further gains in sensitivity and specificity might be achieved at the ≥3 (*moderately stressful*) and 5 (*extremely stressful*) cut-offs. As shown in Table 2, a ≥3 cut-off resulted in sensitivity exceeding 90% in six of the seven occupational samples. However, the increase in sensitivity was at the cost of specificity, which fell to as low as 24% and failed to exceed 50% in all but one of the samples. A ≥5 cut-off produced the opposite effect: across the samples, sensitivity was exceptionally low and specificity very high reaching no less than 96%. Therefore, our results indicated that the cut-off score of 4 on the SIJSM provided the optimal balance of specificity and sensitivity.

[insert Table 2 about here]

**Comparison of 5- and 9-point Rating Scale**

In analyses comparing the discriminatory power of Smith et al.'s (2000) original 5-point rating scale to that of an extended 9-point scale, the AUC was highest on the 5-point rating scale at the ≥4 cut-off (.71). The AUC was highest at the ≥6 cut-off on the 9-point rating scale (.70). At these cut-offs sensitivity and specificity were almost identical across the two rating scales. There appeared to be no gains in sensitivity and specificity from using the 9-point rating scale. Discriminatory statistics for the two rating scales are displayed in Table 3.

[insert Table 3 about here]

## Discussion

The primary objective of this study was to test the ability of a single-item job stressfulness measure (SIJSM) to discriminate between cases and non-cases of common mental disorder (CMD) indicative of likely minor psychiatric morbidity. Discriminatory

power varied across the occupational groups in our analyses with sensitivity and specificity exceeding 70% in the two samples where at least 50% of respondents reported high job stressfulness (those reporting ≥4, their job being *very stressful* or *extremely stressful*). In these two samples the prevalence of CMD was consistent with that of the other participant samples, suggesting that it was the prevalence of high job stressfulness rather than the extent of CMD that accounted for high sensitivity and specificity. These findings suggest that in high stress occupational groups, defined as those where at least half of a workgroup reports high job stressfulness, the SIJSM can be an effective tool to efficiently and non-intrusively identify groups at risk of minor psychiatric morbidity. These groups can then be followed up with further in-depth psychosocial risk assessment that informs targeted interventions.

Locating the cut-off for the identification of high job stressfulness at the third point (*moderately stressful*) on the 5-point scale resulted in exceptionally high sensitivity – more than 90% in six of the seven occupational samples – but very low specificity. This suggests a high true positive and low false negative rate, with most GHQ-12 cases experiencing their job as at least moderately stressful. The proportion of GHQ-12 non-cases experiencing their job as not at all stressful or mildly stressful was small, generating a low true negative rate and high false positive rate, with reports on degree of job stressfulness distributed across the five response categories. Conversely, placement of the cut-off at the fifth (*extremely stressful*) point had the opposite effect, with very low sensitivity relative to specificity of between 96% and 100%. In this scenario, only a small proportion of GHQ-12 cases and almost no GHQ-12 non-cases reported their job as being extremely stressful. The ≥4 (*very stressful* and *extremely stressful*) cut-off offers appears optimal for most applications including organizational psychosocial risk assessment, where there is an imperative to maximize both sensitivity and specificity. In our study, this cut-off resulted in sensitivity and specificity ≥70% in two samples. Echoing Hosmer and Lemeshow's (2000) recommendation that an

AUC of .70 to .79 indicates acceptable discrimination, the SIJSM might be most usefully applied in organizational psychosocial risk assessment activities with occupational groups where sensitivity and specificity have been empirically demonstrated as meeting a ≥70% acceptability threshold. Yet in circumstances where an imperative exists to maximize the true positive rate (i.e., maximize the proportion of CMD cases correctly identified by the SIJSM while minimizing false-negative results) the ≥3 cut-off may be preferable. Such would be the case where the priority is to identify as many CMD cases as possible, i.e., where there were high costs associated with false negatives, plentiful resources for additional and more detailed assessment, and where such assessment would not be too demanding or disruptive. Contrastingly, in circumstances where the true negative rate is of paramount importance (i.e., the focus is on maximizing the proportion of correctly identified CMD non-cases and minimizing false-positive results), the ≥5 cut-off may be more suitable. This approach would be effective when there is a need to identify the absence rather than presence of minor psychiatric morbidity quickly and reliably. This may be the imperative in a large population that typically experiences low prevalence of CMD.

Based on the original ≥4 (*very stressful* and *extremely stressful*) cut-off, two high stress job roles were identified in our study: prison officer and public protection unit police officer. The prevalence of jobs within a nation's workforce that fulfil the definition of high stress whereby at least half of a workgroup reports high job stressfulness is unclear; British nationally representative workforce surveys using the same SIJSM consistently produce a far lower overall rate of 12-19% and do not present results stratified by occupation (Health and Safety Executive, 2012; Houdmont et al., 2010; Scottish Government, 2018). Further testing of the SIJSM with a variety of high stress occupational groups would usefully establish the extent to which its comparatively high discriminatory power identified in our prison officer and public protection unit police officer samples generalises to other high stress occupations

such as, among others, firefighters (Payne & Kinman, 2019) and social workers (Kinman & Grant, 2020). Such corroboration would lend support to the proposition that the SIJSM has particular utility, supported by acceptable sensitivity and specificity (both ≥70%), in high stress occupations.

Our findings appear to suggest that sensitivity and specificity of the SIJSM is related to the prevalence of high job stressfulness; in occupational groups where prevalence of high job stressfulness was ≥50% the SIJSM displayed both sensitivity and specificity exceeding ≥70%. Other explanations for the measure's acceptable sensitivity and specificity with these occupational groups are nevertheless possible. For instance, particular characteristics of the work of prison officers and public protection police officers, or those who seek and retain employment in these roles and who complete the survey, might be contributory factors. Further research is required to test these possibilities.

The sensitivity and specificity of Smith et al.'s (2000) 5-point rating scale was not improved upon by a 9-point scale. It is possible that the five labelled response options in the shorter scale are entirely sufficient for workers to quantify their job stressfulness meaningfully and accurately. Such a conclusion is in line with studies concerning the validity of single-item measures of constructs such as self-esteem which have found that an extended scale offers no improvements over a 5-point scale (Robins et al., 2001). However, it is possible that certain design features of our 9-point scale limited its ability to provide enhanced discriminatory power. Specifically, the labels on our extended 9-point rating scale were the same as those on the 5-point rating scale. This meant that whereas each response category on the 5-point scale had a label, on the 9-point scale the five labels were distributed evenly across the categories resulting in the second, fourth, sixth, and eighth category not having a label. We formatted the rating scale in this way to ensure consistency of labelling across the two scales. However, it is possible that the absence of labels at alternate points on

the extended scale made it less interpretable (Weijeters et al., 2010), undermining any facilitation of respondents' finer grained judgements on their degree of global job stressfulness. Further research involving different versions of an extended scale involving, for example, a label on every response category versus labels on endpoints only and differing numbers of response categories, would usefully confirm whether the provision of additional response categories can enhance discriminatory power. Some single-item measures take a multi-faceted approach to facilitating discrimination between response options. For instance, the Defense and Veterans Pain Rating Scale (DVPRS) (Polomano et al., 2016), which was developed to address shortcomings and inconsistencies in earlier pain rating scales, involves 11 scale points, each of which has a functional pain descriptor (e.g., 'sometimes distracts me') that is supplemented by a traffic-light colour coding system (green for mild pain, yellow for moderate pain, and red for severe pain) and pictorial facial expressions.

Stability of the SIJSM is also an important prerequisite for its regular and repeated administration to gain a snapshot of current job stressfulness and identify temporal changes in CMD. Arapovic-Johansson et al. (2017) concluded that the test-retest reliability of a single-item measure of stress that was non-domain-specific (i.e., did not stipulate that responses should consider specific life domains such as work or domestic life) and focused on mental states, made it suitable for the continual monitoring of stress within groups of workers. Our findings indicate that the SIJSM used in our study could be further tested to examine whether responses fluctuate in line with fluctuations in CMD. If so, the SIJSM would represent a useful practical method for tracking workforce mental wellbeing and the effects of stress management interventions on the prevalence of self-reported CMD.

Across our seven samples, two occupational groups reported prevalence of high job stressfulness of at least 50%, four reported prevalence between 30% and 39% and one group reported a prevalence rate below 20%. While we were able to show sensitivity and specificity

exceeded 70% in the samples where at least 50% of respondents reported high job stressfulness, our data did not allow for the examination of the ability of the SIJSM to identify CMD cases where 40-49% of a workgroup report high job stressfulness. As such, it is not clear what the minimum proportion of a workgroup reporting high job stressfulness might need to be for sensitivity and specificity to remain at or above 70%. Further research with different occupational samples would offer clarity in this regard.

The GHQ-12 is widely used in occupational and general population settings to assess mental wellbeing. We used this instrument as the validation criterion because of its receptiveness to administration within a self-report survey, minimal assessment burden, and the existence of a threshold score that permits classification of respondents as cases and non-cases of probable minor psychiatric disorder. Though extensive validation work has shown the GHQ-12 to be an effective screener for probable minor psychiatric disorder when tested against structured psychiatric interview data, the optimal threshold varies across populations and the AUC is always imperfect at less than 1.00 (e.g., Aalto et al., 2012; Donath, 2001; Lundin et al., 2016). Moreover, UK research involving GHQ-12 validation against clinical interview data in primary care patients suggests being in employment is associated with increased odds of identification as a false positive case (Bell et al., 2005). Taken together, these findings suggest that the GHQ-12 falls short of the standard offered by a structured psychiatric interview. To address these limitations future research should involve interview-based diagnosis involving the widely used Composite International Diagnostic Interview (CIDI) as the validation criterion as well as the development of sector- and role-specific threshold scores.

In conclusion, our findings suggest that a single-item measure of global job stressfulness may be used to classify cases and non-cases of self-reported CMD with an acceptable degree of accuracy in high stress occupational groups, defined as those where at

least 50% of respondents report high job stressfulness (the job being *very stressful* or *extremely stressful*). These findings suggest that the measure may represent an effective tool for the efficient identification of likely cases of minor psychiatric disorder in high stress occupational groups.

## References

Aalto, A-M., Elovainio, M., Kivimäki, M., Uutela, A., & Pirkola, S. (2012). The Beck Depression Inventory and General Health Questionnaire as measures of depression in the general population: A validation study using the Composite International Diagnostic Interview as the gold standard. *Psychiatry Research, 197*(1-2), 163-171. https://doi.org/10.1016/j.psychres.2011.09.008

Addley, K., Toner, A., Erskine, R., & Duffy, U. (2006). *Northern Ireland Civil Service workforce health and well-being survey 2005*. Northern Ireland Statistics and Research Agency.

Allisey, A. F., Noblet, A. J., LaMontage, A. D., & Houdmont, J. (2014). Testing a model of officer intentions to quit: The mediating effects of job stress and job satisfaction. *Criminal Justice and Behavior, 41*(6), 751–771. https://doi.org/10.1177/0093854813509987

Arapovic-Johansson, B., Wåhlin, C., Hagberg, J., Kwak, L., Axén, I., Björklund, C., & Jensen, I. (2020). Experience of stress assessed by text messages and its association with objective workload – a longitudinal study. *International Journal of Environmental Research and Public Health, 17*(3), Article 680. https://doi.org/10.3390/ijerph17030680

Arapovic-Johansson, B., Wåhlin, C., Kwak, L., Björklund, C., & Jensen, I. (2017). Work-related stress assessed by a text message single-item question. *Occupational Medicine, 67*(8), 601-608. https://doi.org/10.1093/occmed/kqx111

Bell, T., Watson, M., Sharp, D., Lyons, I., & Lewis, G. (2005). Factors associated with being a false positive on the General Health Questionnaire. *Social Psychiatry and Psychiatric Epidemiology, 40*(5), 402-407. https://doi.org/10.1007/s00127-005-0881-6

Bergkvist, L., & Rossiter, J. (2007). The predictive validity of multiple-item versus single-item measures of the same constructs. *Journal of Marketing Research, 44*(2), 175-184. https://doi.org/10.1509%2Fjmkr.44.2.175

Blake, H., Yildirim, M., Wood, B., Knowles, S., Mancini, H., Coyne, E., & Cooper, J. (2020). COVID-Well: Evaluation of the implementation of supported wellbeing centres for hospital employees during the COVID-19 pandemic. *International Journal of Environmental Research and Public Health, 17*(24), Article 9401. https://doi.org/10.3390/ijerph17249401

Brough, P., & Biggs, A. (2015). Job demands x job control interaction effects: Do occupation-specific job demands increase their occurrence? *Stress and Health, 31*, 138-149. https://doi.org/10.1002/smi.2537

Cohen, J. (1988). *Statistical power and analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.

Dawes, J. (2008). Do data characteristics change according to the number of scale points used? An experiment using 5-point, 7-point and 10-point scales. *International Journal of Market Research, 50*(1), 61-77. https://doi.org/10.1177/147078530805000106

Donath, S. (2001). The validity of the 12-item General Health Questionnaire in Australia: A comparison between three scoring methods. *Australian and New Zealand Journal of Psychiatry, 35*(2), 231-235. https://doi.org/10.1046%2Fj.1440-1614.2001.00869.x

Elliott-Davies, M. (2018). *Police Federation of England and Wales Demand, Capacity, and Welfare Survey 2018: Headline statistics*. https://www.polfed.org/media/14060/demandcapacityandwelfaresurveyheadlinestatistics2018-06-02-19-v1.pdf

Elliott-Davies, M. (2021). *Police Federation of England and Wales Demand, Capacity, and Welfare Survey 2020: Headline statistics*. https://www.polfed.org/media/16555/dcw_prrb-report-13-01-2021-v20.pdf

Elo, A-L., Leppänen, A., & Jahkola, A. (2003). Validity of a single item measure of stress symptoms. *Scandinavian Journal of Work, Environment and Health, 29*(6), 444-451. www.jstor.org/stable/40967322

Eng, A., Mannetje, A. L., Pearce, N., & Douwes, J. (2011). Work-related stress and asthma: Results from a workforce survey in New Zealand. *Journal of Asthma, 48*(8), 783–789. https://doi.org/10.3109/02770903.2011.608459

EuroFound. (2016). *Sixth European Working Conditions Survey – Overview Report*. Publications Office of the European Union. https://www.eurofound.europa.eu/publications/report/2016/working-conditions/sixth-european-working-conditions-survey-overview-report

Fisher, G. G., Matthews, R. A., & Gibbons, A. M. (2016). Developing and investigating the use of single-item measures in organizational research. *Journal of Occupational Health Psychology, 21*(1), 3–23. https://doi.org/10.1037/a0039139

Gilbert, S., & Kelloway, E. K. (2014). Using single items to measure job stressors. *International Journal of Workplace Health Management, 7*(3), 186-199. https://doi.org/10.1108/ijwhm-03-2013-0011

Goggin, S., & Stoker, L. (2014, September 28-31). *Optimal scale length and single-item attitude measures: Evidence from simulations and a two-wave experiment* [Paper presentation]. American Political Science Association 2014 Annual Meeting, Washington, DC, United States. https://ssrn.com/abstract=2455794

Goldberg, D. P., Gater, R., Sartorius, N., Ustun, T. B., Piccinelli, M., Gureje, O., & Rutter, C. (1997). The validity of two versions of the GHQ in the WHO study of mental illness

in general health care. *Psychological Medicine 27*(1), 191–197. https://doi.org/10.1017/s0033291796004242

Goldberg, D. P., Oldehinkel, T., & Ormel, J. (1998). Why GHQ threshold varies from one place to another. *Psychological Medicine, 28*(4), 915-921. https://doi.org/10.1017/s0033291798006874

Goldberg, D. P., & and Williams, P. (1988). *User's Guide to the General Health Questionnaire*. NFER-Nelson.

Goodwin, L., Ben-Zion, I., Fear, N. T., Hotopf, M., Stansfeld, S. A., & Wessely, S. (2013). Are reports of psychological stress higher in occupational studies? A systematic review across occupational and population based studies. *PLoS ONE*, 8*(11),* Article e78693. https://doi.org/10.1371/journal.pone.0078693

Hardy, G. E., Shapiro, D. A., Haynes, C. E., & Rick, J. E. (1999). Validation of the General Health Questionnaire-12: Using a sample of employees from England's health care services. *Psychological Assessment, 11*(2), 159-165. https://doi.org/10.1037/1040-3590.11.2.159

Health and Safety Executive. (2012). *Psychosocial working conditions in Britain in 2010*. http://www.hse.gov.uk/statistics/pdf/pwc2010.pdf.

Hodgson, J. T., Jones, J. R., Clarke, S. D., Blackburn, A. J, Webster, S., Huxtable, C. S., & Wilkinson, S. (2006). *Workplace health and safety survey programme: 2005 worker survey first findings report*. Health and Safety Executive. https://www.hse.gov.uk/statistics/pdf/whassw1.pdf

Hosmer, D. W., & Lemeshow, S. (Eds.) (2000). *Applied logistic regression* (2nd ed.). John Wiley and Sons.

Houdmont, J. (2014). *Custody officers' stress-related working conditions: Relations with health and organisational effectiveness*. [Unpublished research report for the

Sergeants' Central Committee of the Police Federation of England and Wales]. School of Medicine, University of Nottingham.

Houdmont, J., Cox, T., & Griffiths, A. (2010). Work-related stress case definitions and prevalence rates in national surveys. *Occupational Medicine, 60*(8), 658-661. https://doi.org/10.1093/occmed/kqq138

Houdmont, J., Jachens, L., Randall, R., & Colwell, J. (2021). English rural policing: Job stress and psychological distress. *Policing: An International Journal of Police Strategies and Management, 44*(1), 49-62. https://doi.org/10.1108/PIJPSM-03-2020-0037

Houdmont, J., Jachens, L., Randall, R., Hopson, S., Nuttall, S., & Pamia, S. (2019). What does a single-item measure of job stressfulness assess? *International Journal of Environmental Research and Public Health, 16*(9), Article 1480. https://doi.org/10.3390/ijerph16091480

Houdmont, J., Kerr, R., & Addley, K. (2012). Psychosocial factors and economic recession: The Stormont Study. *Occupational Medicine, 62*(2), 98–104. https://doi.org/10.1093/occmed/kqr216

Houdmont, J., & Randall, R. (2016). Working hours and common mental disorders in English police officers. *Occupational Medicine*, *66*(9), 713-718. https://doi.org/10.1093/occmed/kqw166

Johnson, S. J., Willis, S. M., & Robertson, I. T. (2018). Cross-validation of a short stress measure: ASSET Pulse. *International Journal of Stress Management, 25*(4), 391-400. https://doi.org/10.1037/str0000097

Kinman, G., & Clements, A. (2021). Prison Officer Association survey of work-related wellbeing. https://www.poauk.org.uk/news-events/news-room/posts/2021/january/circ-005-poa-survey-of-work-related-wellbeing/

Kinman, G., & Grant, L. (2020). Emotional demands, compassion and mental health in social workers. *Occupational Medicine, 70*(2), 89-94. https://doi.org/10.1093/occmed/kqw144

Lachance, J-P, Corbière, M., Hains-Monfette, G., & Bernard, P. (2020). *Clearing your mind of work-related stress through moderate-to-vigorous and leisure-time physical activity: What 'dose' it takes?* MedRxiv. https://doi.org/10.1101/2020.05.11.20097931

Littman, A. J., White, E., Satia, J. A., Bowen, D. J., & Kristal, A. R. (2006). Reliability and validity of 2 single-item measures of psychosocial stress. *Epidemiology, 17*(4), 398–403. https://www.jstor.org/stable/20486241

Lundin, A., Hallgren, M., Theobald, H., Hellgren, C., & Torgén, M. (2016). Validity of the 12-item version of the General Health Questionnaire in detecting depression in the general population. *Public Health*, *136*, 66-74. https://doi.org/10.1016/j.puhe.2016.03.005

Payne, N., & Kinman, G. (2019). Job demands, resources and work-related well-being in UK firefighters. *Occupational Medicine, 69*(8-9), 604-609. https://doi.org/10.1093/occmed/kqz167

Phillips, S. J., Sen, D., & McNamee, R. (2008). Risk factors for work-related stress and health in head teachers. *Occupational Medicine, 58*(8), 584–586. https://doi.org/10.1093/occmed/kqn112

Polomano, R. C., Galloway, K. T., Kent, M. L., Brandon-Edwards, H., Kwon, K. N., Morales, C., & Buckenmaier III, C. T. (2016). Psychometric testing of the Defense and Veterans Pain Rating Scale (DVPRS): A new pain scale for military population. *Pain Medicine*, *17*(8), 1505-1519.

Preston, C. C., & Colman, A. M. (2000). Optimal number of response categories in rating scales: reliability, validity, discriminating power, and respondent preferences. *Acta Psychologica, 104*(1), 1-15. https://doi.org/10.1016/s0001-6918(99)00050-5

Robins, R., Hendin, H., & Trzesniewski, K. (2001). Measuring global self-esteem: Construct validation of a single-item measure and the Rosenberg self-esteem scale. *Personality and Social Psychology Bulletin, 27*(2), 151-161. https://doi.org/10.1177/0146167201272002

Scarpello, V., & Campbell, J. P. (1983). Job satisfaction: Are all the parts there? *Personnel Psychology*, *36*(3), 577–600. http://dx.doi.org/10.1111/j.1744-6570.1983.tb02236.x

Scottish Government. (2018). *The Scottish Health Survey 2017 edition – summary*. https://www.gov.scot/publications/scottish-health-survey-2017-summary-key-findings/

Smith, A. (2000). The scale of perceived occupational stress. *Occupational Medicine, 50*(5), 294–298. https://doi.org/10.1093/occmed/50.5.294

Smith, A. (2001). Perceptions of stress at work. *Human Resource Management, 11*(4), 74–86. https://doi.org/10.1111/j.1748-8583.2001.tb00052.x

Smith, A., Johal, S., Wadsworth, E., Davey Smith, G., & Peters, T. (2000). *The scale of perceived stress at work: The Bristol stress and health at work study*. Health and Safety Executive. https://www.hse.gov.uk/research/crr_pdf/2000/crr00265.pdf

Szeto, A. C. H., & Dobson, K. S. (2013). Mental disorders and their association with perceived work stress: An investigation of the 2010 Canadian Community Health Survey. *Journal of Occupational Health Psychology, 18*(2), 191-197. https://doi.org/10.1037/a0031806

Trevethan, R. (2017). Sensitivity, specificity, and predictive values: Foundations, pliabilities, and pitfalls in research and practice. *Frontiers in Public Health, 5*, Article 307. https://doi.org/10.3389/fpubh.2017.00307

Trull, T. J., & Ebner-Priemer, U. W. (2009). Using experience sampling methods/ecological momentary assessment (ESM/EMA) in clinical assessment and clinical research: introduction to the special section. *Psychological Assessment, 21*(4), 457-462. https://doi.org/10.1037/a0017653

Weng, L-J. (2004). Impact of the number of response categories and anchor labels on coefficient alpha and test-retest reliability. *Educational and Psychological Measurement, 64*(6), 956-972. https://doi.org/10.1177/0013164404268674

Wanous, J. P., Reichers, A. E., & Hudy, M. J. (1997). Overall job satisfaction measures: how good are single-item measures? *Journal of Applied Psychology, 82(2)*, 247-252. https://doi.org/10.1037/0021-9010.82.2.247

Weijeters, B., Cabooter, E., & Schillewaert, N. (2010). The effect of rating scale format on response styles: The number of response categories and response category labels. *International Journal of Research in Marketing*, *27*(3), 236-247. https://doi.org/10.1016/j.ijresmar.2010.02.004

Table 1

*Personal Characteristics, High Job Stressfulness and Common Mental Disorder Caseness Prevalence, and Association and Discrimination Statistics (≥4 Cut Off)*

| | Civil servants *n*=16,495 | Prison officers *n*=1,837 | Police officers – public protection *n*=452 | Police officers – rural *n*=260 | Police officers – response *n*=497 | Police officers – crime investigation *n*=215 | Police officers – custody *n*=902 |
|---|---|---|---|---|---|---|---|
| *Personal characteristics* | | | | | | | |
| Gender (% male) | 48.0 | 86.8 | 40.9 | 74.6 | 72.2 | 67.9 | 82.8 |
| Age, *M* (*SD*) | 39.07 (10.79) | 48.30 (9.44) | 40.68 (7.01) | 43.36 (9.38) | 41.14 (7.64) | 43.03 (7.68) | 43.78 (6.45) |
| *Prevalence statistics* | | | | | | | |
| High job stressfulness (%, 95% CI) | 18.4 (17.8-19.0) | 53.1 (50.8-55.4) | 49.6 (44.9-54.3) | 30.4 (24.9-36.4) | 32.6 (28.5-36.9) | 38.6 (32.1-45.5) | 38.8 (35.6-42.1) |
| Common mental disorder (%, 95% CI) | 27.4 (26.8-28.1) | 58.6 (56.3-60.8) | 56.4 (51.7-61.0) | 44.2 (38.1-50.5) | 48.5 (44.0-53.0) | 50.2 (43.4-57.1) | 56.1 (52.8-59.4) |
| *Association statistics* | | | | | | | |
| Correlation (*r,* 95% CI) | .45* (.43-.46) | .54* (.50-.58) | .50* (.42-.58) | .48* (.38-.59) | .58* (.50-.65) | .53* (.42-.65) | .46* (.40-.52) |
| *Discrimination statistics* | | | | | | | |
| Odds ratio (95% CI) | 6.97 (6.41-7.59) | 7.39 (5.99-9.11) | 8.35 (5.43-12.83) | 8.81 (4.72-16.45) | 9.64 (6.07-15.31) | 8.32 (4.36-15.89) | 4.85 (3.58-6.58) |
| Sensitivity (%) | 42.2 | 72.2 | 70.6 | 53.9 | 55.2 | 61.1 | 54.0 |
| Specificity (%) | 90.5 | 74.0 | 77.7 | 88.3 | 88.7 | 84.1 | 80.6 |
| Cases correctly classified (%) | 77.3 | 72.9 | 73.7 | 73.1 | 72.4 | 72.6 | 65.6 |
| Positive predictive value (%) | 62.7 | 79.7 | 80.4 | 78.5 | 82.1 | 79.5 | 78.0 |
| Negative predictive value (%) | 80.5 | 65.3 | 67.1 | 70.7 | 67.8 | 68.2 | 57.8 |
| Positive likelihood ratio (95% CI) | 4.44 (4.16-4.74) | 2.78 (2.45-3.15) | 3.17 (2.41-4.16) | 4.61 (2.86-7.43) | 4.88 (3.03-5.95) | 3.84 (2.42-6.09) | 2.78 (2.22-3.49) |
| Negative likelihood ratio (95% CI) | 0.64 (0.62-0.66) | 0.38 (0.34-0.42) | 0.38 (0.31-0.46) | 0.52 (0.42-0.64) | 0.51 (0.44-0.60) | 0.46 (0.36-0.59) | 0.57 (0.51-0.63) |
| AUC (95% CI) | .73 (.72-.74) | .73 (.71-.76) | .74 (.69-.79) | .71 (.65-.78) | .72 (.67-.77) | .73 (.66-.80) | .67 (.64-.71) |

* $p < .01$.

Table 2

*Discrimination Statistics (≥3 and 5 Cut Off Points)*

| | Civil servants *n*=16,495 | Prison officers *n*=1,837 | Police officers – public protection *n*=452 | Police officers – rural *n*=260 | Police officers – response *n*=497 | Police officers – crime investigation *n*=215 | Police officers – custody *n*=902 |
|---|---|---|---|---|---|---|---|
| ≥3 cut off | | | | | | | |
| High job stressfulness (%) | 53.7 | 85.3 | 86.3 | 75.8 | 80.5 | 82.8 | 81.9 |
| Sensitivity (%) | 77.1 | 94.7 | 95.4 | 92.2 | 94.2 | 93.5 | 91.9 |
| Specificity (%) | 55.1 | 28.0 | 24.4 | 37.2 | 32.4 | 28.0 | 30.8 |
| AUC (95% CI) | .66 (.65-.67) | .61 (.59-.64) | .59 (.54-.65) | .65 (.58-.71) | .63 (.58-.68) | .61 (.53-.68) | .61 (.58-.65) |
| ≥5 cut off | | | | | | | |
| High job stressfulness (%) | 4.2 | 13.7 | 13.3 | 5.0 | 5.8 | 6.5 | 8.5 |
| Sensitivity (%) | 12.1 | 20.3 | 22.0 | 11.3 | 10.8 | 13.0 | 13.2 |
| Specificity (%) | 98.8 | 96.0 | 99.8 | 100 | 99.9 | 100 | 97.5 |
| AUC (95% CI) | .56 (.54-.57) | .58 (.55-.61) | .60 (.55-.65) | .56 (.49-.63) | .55 (.50-.60) | .57 (.49-.64) | .55 (.52-.59) |

Table 3

*5- and 9-Point Rating Scale Discrimination for Common Mental Disorder (n=260)*

| High job stressfulness threshold score | Sensitivity (%) | Specificity (%) | Cases correctly classified (%) | Positive predictive value (%) | Negative predictive value (%) | Positive likelihood ratio | Negative likelihood ratio | AUC (95% CI) |
|---|---|---|---|---|---|---|---|---|
| 5-point rating scale | | | | | | | | |
| 1 (*not at all stressful*) | - | - | - | - | - | - | - | - |
| ≥2 (*mildly stressful*) | 100 | 4.8 | 46.9 | 45.5 | 100 | 1.05 | - | .52 (.45-.59) |
| ≥3 (*moderately stressful*) | 92.2 | 37.2 | 61.5 | 53.8 | 85.7 | 1.47 | .21 | .65 (.58-.71) |
| ≥4 (*very stressful*) | 53.9 | 88.3 | 73.1 | 78.5 | 70.7 | 4.61 | .52 | .71 (.65-.78) |
| 5 (*extremely stressful*) | 11.3 | 100 | 60.8 | 100 | 58.7 | - | .89 | .56 (.49-.63) |
| 9-point rating scale | | | | | | | | |
| 1 (*not at all stressful*) | - | - | - | - | - | - | - | - |
| ≥2 | 100 | 3.4 | 46.2 | 45.1 | 100 | 1.04 | - | .52 (.45-.59) |
| ≥3 (*mildly stressful*) | 99.1 | 9.0 | 48.8 | 46.3 | 92.9 | 1.09 | .10 | .54 (.47-.61) |
| ≥4 | 89.6 | 35.9 | 59.6 | 52.6 | 81.3 | 1.40 | .29 | .63 (.56-.69) |
| ≥5 (*moderately stressful*) | 87.0 | 46.2 | 64.2 | 56.2 | 81.7 | 1.62 | .28 | .67 (.60-.73) |
| ≥6 | 58.3 | 82.1 | 71.5 | 72.0 | 71.3 | 3.26 | .51 | .70 (.64-.77) |
| ≥7 (*very stressful*) | 38.3 | 94.5 | 69.6 | 84.6 | 65.9 | 6.96 | .65 | .66 (.60-.73) |
| ≥8 | 13.9 | 98.6 | 61.2 | 88.9 | 59.1 | 9.93 | .87 | .56 (.49-.63) |
| ≥9 (*extremely stressful*) | 9.6 | 100 | 60.0 | 100 | 58.2 | - | .90 | .55 (.48-.62) |

*Note*. LR+ and LR- not presented where sensitivity or specificity is 100%.