Short communication

Scand J Work Environ Health. 2014;40(4):428–431. doi:10.5271/sjweh.3428

Work ability as prognostic risk marker of disability pension: single-item work ability score versus multi-item work ability index

Corné AM Roelen, MD, PhD,[1, 2, 3] Willem van Rhenen, MD, PhD,[1] Johan W Groothoff, PhD,[2] Jac JL van der Klink, MD, PhD,[2] Jos WR Twisk, PhD,[3] Martijn W Heymans, PhD [3]

Roelen CAM, van Rhenen W, Groothoff JW, van der Klink JJL, Twisk JWR, Heymans MW. Work ability as prognostic risk marker of disability pension: single-item work ability score versus multi-item work ability index. *Scand J Work Environ Health.* 2014;40(4):428–431. doi:10.5271/sjweh.3428

Objectives Work ability predicts future disability pension (DP). A single-item work ability score (WAS) is emerging as a measure for work ability. This study compared single-item WAS with the multi-item work ability index (WAI) in its ability to identify workers at risk of DP.

Methods This prospective cohort study comprised 11 537 male construction workers, who completed the WAI at baseline and reported DP after a mean 2.3 years of follow-up. WAS and WAI were calibrated for DP risk predictions with the Hosmer-Lemeshow (H-L) test and their ability to discriminate between high- and low-risk construction workers was investigated with the area under the receiver operating characteristic curve (AUC).

Results At follow-up, 336 (3%) construction workers reported DP. Both WAS [odds ratio (OR) 0.72, 95% confidence interval (95% CI) 0.66–0.78] and WAI (OR 0.57, 95% CI 0.52–0.63) scores were associated with DP at follow-up. The WAS showed miscalibration (H-L model χ^2=10.60; df=3; P=0.01) and poorly discriminated between high- and low-risk construction workers (AUC 0.67, 95% CI 0.64–0.70). In contrast, calibration (H-L model χ^2=8.20; df=8; P=0.41) and discrimination (AUC 0.78, 95% CI 0.75–0.80) were both adequate for the WAI.

Conclusion Although associated with the risk of future DP, the single-item WAS poorly identified male construction workers at risk of DP. We recommend using the multi-item WAI to screen for risk of DP in occupational health practice.

Key terms construction worker; discriminative ability; prognostic research; ROC curve; sensitivity; specificity; WAI; WAS

Work disability has become a major occupational health problem in developed economies. On average 6% of the working population in OECD countries received disability pension (DP) benefits in 2009 (1). DP recipient rates are highest in Sweden (11%), Norway (10%), Finland (8%), and The Netherlands (8%). OECD countries spend on average 1.9% of their gross domestic product on DP benefits (2). As a result, governments attach importance to sustained work ability throughout working life. Work ability is developing into a more versatile concept and its definition may differ across settings, for example occupational health, insurance medicine, or rehabilitation (3). In occupational health, work ability is primarily determined by the balance between an individual's work demands and resources.

In the 1980s, Finnish occupational clinicians developed the work ability index (WAI) to measure work ability. Several studies have demonstrated that low or declining WAI levels increase the risk of DP (4–6). The WAI is a long instrument and rather difficult to complete, especially when workers do not have a good understanding of it. A user-friendly single-item work ability score (WAS), asking for current work ability in relation to lifetime best, is increasingly being used to

1 ArboNed Occupational Health Service, PO Box 85091, 3508 AB Utrecht, The Netherlands.
2 Department of Health Sciences, University Medical Center Groningen, University of Groningen, PO Box 196, 9700 AD, Groningen, The Netherlands.
3 Department of Health Sciences, VU University, De Boelelaan 1085-1087, 1081 HV Amsterdam, The Netherlands.

Correspondence to: CAM Roelen, ArboNed, PO Box 85091, 3508 AB Utrecht, The Netherlands. [E-mail: corne.roelen@arboned.nl]

assess work ability (7). The similarity in results between WAS and WAI fosters the use of the WAS for large-scale population surveys (8). The present study compares the single-item WAS with the multi-item WAI in its ability to identify workers at risk of DP and discriminate between workers at high and low risk of DP.

Methods

ArboNed is a nationwide occupational health service in The Netherlands. ArboNed recruits workers from contracted construction companies every two years for health checks assessing work ability. All 18 093 construction workers who participated in health checks in the period 2005–2007 were included in this study. After a mean 2.3 (standard deviation [SD] 0.1) years, 11 537 (64%) of these workers participated in another health check, assessing DP with the question "At the moment, are you receiving disability pension benefits?" (no/yes). The Medical Ethics Committee of the University Medical Center Groningen approved the study.

Work ability

The single-item WAS asked construction workers "Assume that your work ability at its best has a value of 10 points. How many points would you give your current work ability?" Workers could respond on a 10-point scale, ranging from 0 (completely unable to work) to 10 (work ability at its best).

The WAI is a self-administered instrument that asks for current work ability (as the WAS does), work ability in relation to physical and mental job demands, and work ability in the forthcoming two years. The WAI assesses diseases by a list of medical diagnoses. For this study, we used a short version with 15 medical diagnoses (9). Furthermore, the WAI asks for impaired work performance due to illness and sickness absence over the last 12 months. Finally, mental resources are addressed with the items: "Have you been able to enjoy your regular daily activities?", "Have you been active and alert?", and "Have you felt yourself to be full of hope about the future?". All WAI items are weighted and summed to a composite score ranging from 7=poor work ability to 49=excellent work ability (10).

Statistical analysis

Statistical analyses were performed in SPSS for Windows, version 20 (IBM Corp, Armonk, NY, USA). WAS and WAI scores were standardized as percentage of their maximum score and included as continuous independent variables in separate logistic regression models with DP (no=0, yes=1) at follow-up as outcome variable. Odds ratios (OR) and related 95% confidence intervals (95% CI) are presented per 10-point increase in standardized WAS and WAI scores, adjusted for age, working hours/week, and number of years employed in the construction industry.

The predictive performance of both logistic regression models was quantified in terms of calibration and discrimination. Calibration refers to the agreement between predicted and observed DP risks, and was investigated by the Hosmer-Lemeshow (H-L) test. Lower H-L model χ^2 indicates better calibration and H-L $P \geq 0.05$ reflects adequate calibration (11). Discrimination refers to the ability to distinguish between workers at high and low risk of DP, and was examined by receiver operating characteristic (ROC) analysis. An area under the ROC curve (AUC) ≥ 0.75 represents adequate discrimination (11, 12).

Results

Men participating in both health checks were aged 45.4 (SD 9.5) years and those who did not participate in the health check at follow-up were aged 43.1 (SD 11.1 years; t-test $P<0.01$) years. Participants had been working 39.7 (SD 7.3) hours/week for on average 24.4 (SD 12.1) years in the construction industry, while men who did not participate at follow-up worked 39.4 (SD 7.9; t-test $P=0.67$) hours/week for on average 23.8 (SD 11.7; t-test $P=0.29$) years.

Six construction workers had missing data on the WAS and 2007 (17%) workers had missing data on the WAI. The work ability scores did not differ between men who did (WAS 7.8, SD 1.3; WAI 40.1, SD 4.9) and did not (WAS 7.9, SD 1.3, t-test $P=0.44$; WAI 41.3, SD 5.1, t-test $P=0.17$) participate at follow-up.

At baseline, 81% were blue-collar workers versus 80% at follow-up. WAS scores did not differ between blue- (mean 7.8, SD 1.4) and white- (mean 8.1, SD 1.3; t-test $P=0.63$) collar workers, whereas WAI scores were lower among blue- (mean 39.8, SD=5.0) than white- (mean 41.3, SD=4.4; t-test $P<0.01$) collar workers.

A total of 336 (3%) construction workers [287 (4%) blue- and 49 (3%) white-collar workers] reported DP at follow-up. Both standardized WAS (OR 0.72, 95% CI 0.66–0.78) and WAI (OR 0.57, 95% CI 0.52–0.63) scores were associated with DP. Associations did not differ significantly between blue- and white-collar workers [data not shown].

The WAS showed miscalibration (H-L model $\chi^2=10.60$; df=3; $P=0.01$) indicating that it did not accurately predict the risk of DP. Figure 1 shows that the WAS poorly discriminated (AUC 0.67, 95% CI 0.64–0.70) between workers at high and low risk of DP. This was

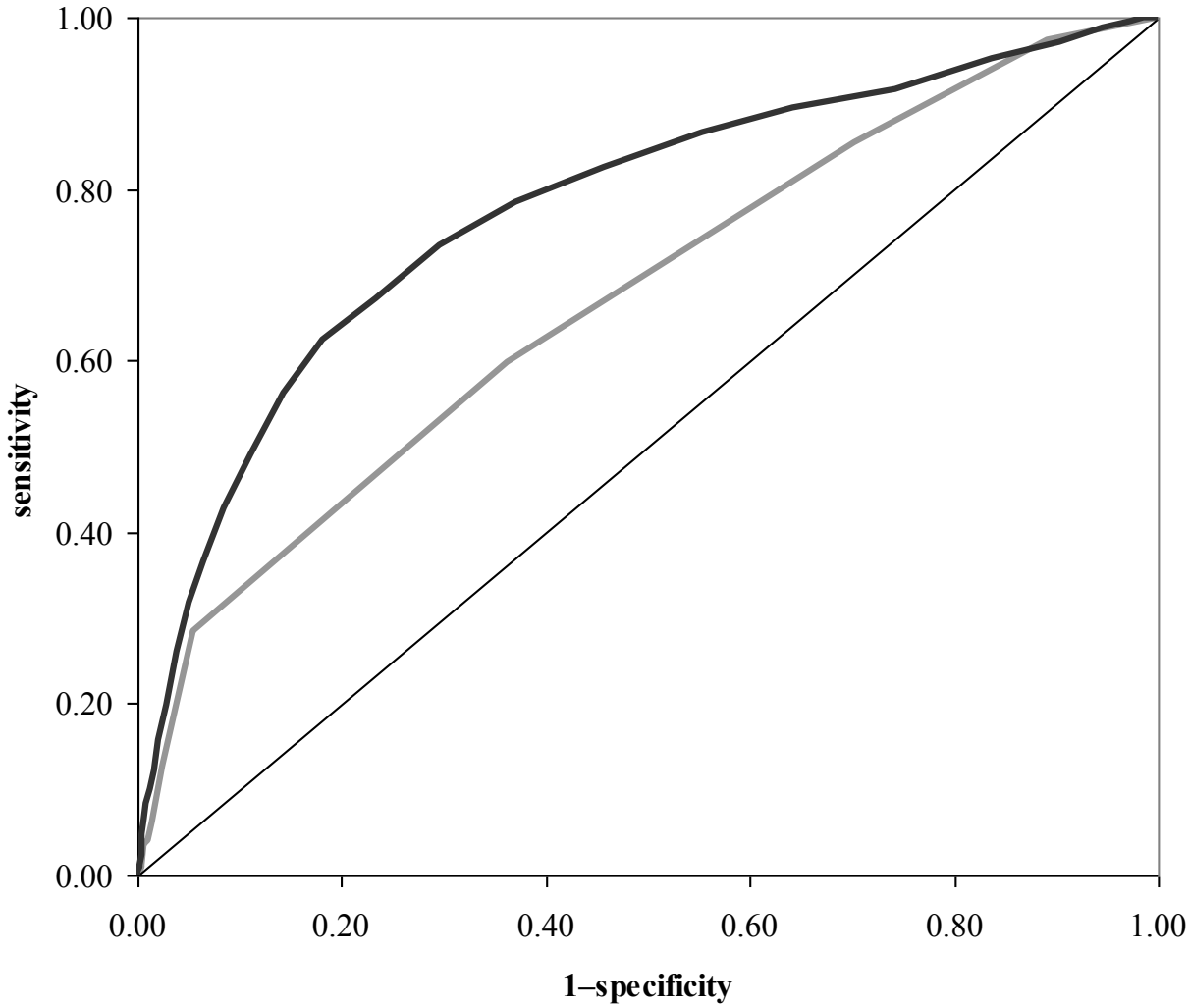

Figure 1. The figure shows the ability of the single-item work ability score [(WAS) grey line with area under the curve (AUC) 0.67] and the multi-item work ability index [(WAI) black line with AUC 0.78] to discriminate construction workers at high risk of disability pension from those at low risk; the diagonal indicates no discrimination above chance.

Table 1. Sensitivities and specificities at different work ability cut-off scores.

Work ability score (N=11 531)			Work ability index (N=9350)		
Score	Sensitivity	Specificity	Score	Sensitivity	Specificity
1	0.01	1.00	34	0.43	0.91
2	0.01	1.00	35	0.49	0.89
3	0.04	0.99	36	0.56	0.86
4	0.04	0.99	37	0.63	0.82
5	0.06	0.99	38	0.67	0.77
6	0.13	0.97	39	0.74	0.70
7	0.29	0.95	40	0.79	0.63
8	0.60	0.64	41	0.82	0.54
9	0.85	0.30	42	0.87	0.45
10	0.97	0.11	43	0.90	0.36

probably due to the fact that 94% of workers scored WAS ≥7. Specificities were high, but sensitivities low for WAS <7 (table 1).

For the multi-item WAI, calibration (H-L model χ^2=8.20; df=8; P=0.41) and discrimination (AUC 0.78, 95% CI 0.75–0.80) were both adequate. WAI scores of 37–40 were the most optimal cut-offs (table 1).

Discussion

The multi-item WAI, but not the single-item WAS showed adequate calibration and discrimination to identify construction workers at increased risk of DP. It should be acknowledged that our results only apply

to male construction workers and may be different for women and workers in other economic sectors. Ahlstrom et al (7) reported that the WAS is a good alternative to the WAI for assessing the status and progress of work ability among sick-listed female human service workers. In line with our present results, the authors discussed that the WAI better predicted future health outcomes. El Fassi et al (8) demonstrated that the WAS collects information on work ability as validly as the WAI. Despite being a valid measure for work ability, our results show that the ability of the WAS to identify construction workers at increased risk of DP is poor.

Practical implications

The WAS discriminates to some extent between male construction workers at high and low risk of DP. When we want to predict a multi-factorial endpoint such as DP, an AUC of 0.67 might not be that bad. However, the multi-item WAI better discriminates between high- and low-risk workers. Critics argue that the WAI is a long and complicated instrument to complete, which might explain the 17% of missing responses on the WAI in our study. The WAS is more user-friendly and easier to interpret. Furthermore, the WAS can be implemented at lower cost in large-scale surveys. We could consider the WAS as a primary screening instrument and then distribute the WAI only to the workers with low WAS scores. The present study showed low sensitivities for WAS cut-off scores <8, but for WAS ≥9 sensitivity was acceptable. In subgroup analysis, we found that the discriminative ability of the WAI was lower among workers with WAS <9 scores than in the total population of male construction workers [data not shown]. This may be due to the reduced spread of work ability scores among the selected workers or the so-called "incorporation bias" as the WAS is part of the WAI. Based on this finding, we advise against using the WAS as a primary screening instrument and recommend the short version of the WAI listing 15 medical diagnoses to screen construction workers for risk of DP.

In a random sample of the German workforce, Bethge et al (13) found that WAI ≤37 identified workers in need of rehabilitation services. We recommend WAI cut-off scores of 37–40 to identify construction workers at increased risk of DP. At a cut-off point <36, about half of DP cases would be missed, while at cut-off points >40 about half of the workers are falsely identified as being at increased risk of DP. Which cut-off point between 37 and 40 should be chosen depends on the burden and costs of interventions to prevent DP.

References

1. Organization for Economic Cooperation and Development. Sickness, disability and work: breaking the barriers. Paris: OECD Publishing; 2010.
2. Organization for Economic Cooperation and Development. Society at a Glance 2011: OECD Social Indicators. Paris: OECD Publishing; 2011.
3. Ilmarinen J. Work ability – a comprehensive concept for occupational health research and prevention. Scand J Work Environ Health 2009;35:1-5. http://dx.doi.org/10.5271/sjweh.1304
4. Liira J, Matikainen E, Leino-Arjas P, Malmivaara A, Mutanen P, Rytkönen H, Juntunen J. Work ability of middle-aged Finnish construction workers – a follow-up study in 1991–1995. Int J Ind Ergon. 2000;25:477-81. http://dx.doi.org/10.1016/S0169-8141(99)00032-3
5. Alavinia SM, de Boer AGEM, van Duivenbooden JC, Frings-Dresen MH, Burdorf A. Determinants of work ability and its predictive value for disability. Occup Med. 2009; 59:32-7. http://dx.doi.org/10.1093/occmed/kqn148
6. Bethge M, Gutenbrunner C, Neuderth S. Work Ability Index predicts application for disability pension after work-related medical rehabilitation fro chronic back pain. Arch Phys Med Rehabil. 2013;94:2262-8. http://dx.doi.org/10.1016/j.apmr.2013.05.003
7. Ahlstrom L, Grimby-Ekman A, Hagberg M, Dellve L. The work ability index and single-item question: associations with sick leave, symptoms, and health – a prospective study of women on long-term sick leave. Scand J Work Environ Health. 2010;36:404-12. http://dx.doi.org/10.5271/sjweh.2917
8. El Fassi M, Bocquet V, Majeri N, Lair ML, Couffignal S, Mairiaux P. Work ability assessment in a worker population: comparison and determinants of Work Ability Index and Work Ability score. BMC Public Health. 2013;13:305. http://dx.doi.org/10.1186/1471-2458-13-305
9. Tuomi K, Ilmarinen J, Jahkola A, Katajarinne L, Tulkki A. Work ability index, second revised version. Helsinki: Finnish Institute of Occupational Health; 1998.
10. Steyerberg EW. Clinical prediction models. Springer: New York; 2009. http://dx.doi.org/10.1007/978-0-387-77244-8
11. Fad J, Upadhye S, Worster A. Understanding receiver operating characteristic (ROC) curves. CJEM 2006; 8:19-20.
12. Bethge M, Radoschewski FM, Gutenbrunner C. The work ability index as a screening tool to identify the need for rehabilitation: longitudinal findings from the second German sociomedical panel of employees. J Rehabil Med. 2012;44:980-7. http://dx.doi.org/10.2340/16501977-1063

Received for publication: 29 November 2013