

Validating abbreviated measures of effort-reward imbalance at work in European cohort studies: the IPD-Work consortium

Johannes Siegrist · Nico Dragano · Solja T. Nyberg · Thorsten Lunau · Lars Alfredsson · Raimund Erbel · Göran Fahlén · Marcel Goldberg · Karl-Heinz Jöckel · Anders Knutsson · Constanze Leineweber · Linda L. Magnusson Hanson · Maria Nordin · Reiner Rugulies · Jürgen Schupp · Archana Singh-Manoux · Töres Theorell · Gert G. Wagner · Hugo Westerlund · Marie Zins · Katriina Heikkilä · Eleonor I. Fransson · Mika Kivimäki

Received: 13 July 2012 / Accepted: 7 February 2013
© Springer-Verlag Berlin Heidelberg 2013

Abstract

Background Effort-reward imbalance (ERI) is an established conceptualisation of work stress. Although a validated effort-reward questionnaire is available for public use, many epidemiological studies adopt shortened scales and proxy measures. To examine the agreement between different abbreviated measures and the original instrument, we compared different versions of the effort-reward scales available in 15 European cohort studies participating in the IPD-Work (Individual-participant-data meta-analysis in working populations) consortium.

This study was conducted for the IPD-Work consortium.

Johannes Siegrist and Nico Dragano: Equal contribution as first authors.

Electronic supplementary material The online version of this article (doi:10.1007/s00420-013-0855-z) contains supplementary material, which is available to authorized users.

J. Siegrist · N. Dragano (✉) · T. Lunau
Department of Medical Sociology, Medical Faculty, University of Düsseldorf, Düsseldorf, Germany
e-mail: dragano@med.uni-duesseldorf.de

S. T. Nyberg · K. Heikkilä · M. Kivimäki
Finnish Institute of Occupational Health, Helsinki, Finland

L. Alfredsson · E. I. Fransson
Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden

R. Erbel
Department of Cardiology, West-German Heart Center Essen, University Duisburg-Essen, Essen, Germany

G. Fahlén
The National Agency for Special Needs Education and Schools, Härnösand, Sweden

Methods Five of the 15 studies provide information on the original ('complete') scales measuring 'effort' and 'reward', whereas the 10 remaining studies used 'partial' scales. To compare different versions of the ERI scales, we analyse individual-level data from 31,790 participants from the five studies with complete scales.

Results Pearson's correlation between partial and complete scales was very high in case of 'effort' (where 2 out of 3 items were used) and very high or high in case of 'reward', if at least 4 items (out of 7) were included. Reward scales composed of 3 items revealed good to satisfactory agreement, and in one case, a reward scale consisting of 2 items only demonstrated a modest, but still acceptable degree of agreement. Sensitivity and specificity of a composite measure, the ratio of effort and reward, comparing partial versus complete scales ranged between 59–93 and 85–99 %, respectively. Complete and partial scales were strongly associated with poor self-rated health.

M. Goldberg · M. Zins
Versailles-Saint Quentin University, Versailles, France

M. Goldberg · A. Singh-Manoux · M. Zins
Inserm U1018, Centre for Research in Epidemiology and Population Health, Villejuif, France

K.-H. Jöckel
Institute for Medical Informatics, Biometry, and Epidemiology, University Duisburg-Essen, Essen, Germany

A. Knutsson
Department of Health Sciences, Mid Sweden University, Sundsvall, Sweden

C. Leineweber · L. L. Magnusson Hanson · T. Theorell · H. Westerlund · E. I. Fransson
Stress Research Institute, Stockholm University, Stockholm, Sweden

Conclusion Our results support the notion that short proxy measures or partial versions of the original scales can be used to assess effort-reward imbalance.

Keywords Effort-reward imbalance · Work stress · IPD-Work · Abbreviated measures · European cohort studies

Introduction

Chronic psychosocial stress at work has been identified as a risk factor of several major non-communicable diseases, most notably cardiovascular disease and depression. Work stress was shown to influence the onset as well as disease prognosis (Schnall et al. 2009; Stansfeld and Candy 2006; Tsutsumi and Kawakami 2004; Nieuwenhuijsen et al. 2010; Steptoe and Kivimäki 2012), with some noticeable exceptions (Belkic et al. 2004). Importantly, work stress also contributes to poor self-rated health (Salavecz et al. 2010), functional limitations (Kuper et al. 2002), sick leave (Head et al. 2007), and disability (Blekesaune and Solem 2005; Dragano 2007). To further improve understanding of stress-related disease, existing evidence from small and medium size studies must be approved using large-scale samples which allow the calculation of precise estimators of effect, subgroup analyses and associations with outcomes of smaller incidence. One promising strategy is the pooling of data from available studies in order to improve statistical power and precision of prediction of work stress-related morbidity and mortality risks. Such an approach requires comparable measures of exposure across studies which are based on theoretical concepts of stress at work.

While several such theoretical concepts have been developed (Cartwright and Cooper 2009), the following theoretical models are often applied in prospective epidemiological studies: the demand-control (DC) (or job strain) model (Karasek and Theorell 1990), focusing on distinct job task profiles, the model of organisational justice

(Greenberg 2010; Elovainio et al. 2002), dealing with fairness of procedures and interactions, and the effort-reward imbalance model (ERI) (Siegrist 1996) which provides the basis of the current report. The ERI model addresses the work contract and the principle of social reciprocity lying at its core. Rewards received in return to efforts spent at work include money, esteem, and career opportunities (promotion, job security). The model proposes that lack of reciprocity (high effort in combination with low reward) generates negative emotions and psychobiological stress responses with adverse long-term effects on health. This model is usually measured by a psychometrically validated questionnaire available in two highly comparable versions, an original version containing 23 Likert-scaled items and a short version containing 16 items measuring the two theoretical dimensions ‘effort’, and ‘reward’, and an additional personal factor ‘over-commitment’, in terms of uni-dimensional scales (Siegrist et al. 2004, 2009; Leineweber et al. 2010; Li et al. 2012a, b; Magnavita et al. 2012).

While a uniform application of these measures in epidemiological investigations is desirable, financial and time constraints often result in the use of even shorter sets of questions, thus precluding an appropriate test of the model by applying abbreviated or proxy measures. Nevertheless, it is worthwhile to compare the psychometric properties of these latter measures with the original measurement. If an acceptable agreement between these different measures is observed, it would be reasonable to pool data from available studies in order to improve statistical power and precision of prediction of work stress-related morbidity and mortality risks.

In this article, we make an inventory of different operational measurements of the ERI model in 15 prospective European cohort studies. In order to assess the quality of measurements based on partial scales, we use data from 5 studies which contained the original ERI scales to test the validity of partial proxy measures by comparing these scales with the original scales.

M. Nordin
Department of Public Health and Clinical Medicine,
Occupational and Environmental Medicine, Umeå University,
Umeå, Sweden

R. Rugulies
National Research Centre for the Working Environment,
Copenhagen, Denmark

R. Rugulies
Department of Public Health, University of Copenhagen,
Copenhagen, Denmark

R. Rugulies
Department of Psychology, University of Copenhagen,
Copenhagen, Denmark

J. Schupp · G. G. Wagner
German Institute for Economic Research, Berlin, Germany

A. Singh-Manoux · M. Kivimäki
Department of Epidemiology and Public Health, University
College London, London, UK

G. G. Wagner
Max Planck Institute for Human Development, Berlin, Germany

E. I. Fransson
School of Health Sciences, Jönköping University,
Jönköping, Sweden

Methods

Study context

This study is part of the IPD-Work (Individual-participant-data meta-analysis in working populations) consortium of European cohort studies (Fransson et al. 2012; Nyberg et al. 2012). The aim of the consortium is to establish a large-scale knowledge base for advanced research on adverse health effects of work stress. To this means, data from available European cohort studies will be harmonised and jointly analysed. The task of comparing the full version of the questionnaire to shorter versions has already been accomplished with regard to the DC model (Fransson et al. 2012). In this article, we test the comparability of different measurements of the ERI model.

Study population and assessments

To assess the concordance between different scales measuring the ERI model, we examined the agreement between the full original scales that were available from 5 studies and the partial proxy scales applied in the remaining 10 investigations (a short description of these studies is given in *ESM Appendix 1*). The 5 studies with original scales were HNR (Heinz Nixdorf Recall Study, Germany), SOEP (German Socio-Economic Panel Study), SLOSH (Swedish Longitudinal Occupational Survey of Health, Sweden), GAZEL (GAZ and ELectricité Study, France), and WOLF-F (Work, Lipids and Fibrinogen, Norrland Follow-up study, Sweden). Response rates of those studies range from 56 % (HNR) to 73 % (GAZEL).

Partial proxy scales were available from BELSTRESS, Belgium, COPSOQ-2 (Copenhagen Psychosocial Questionnaire Study), DWECS (Danish Work Environment Cohort Study), IPAW (Intervention Project on Absence and Well-being Study) and PUMA (Copenhagen Burnout Inventory Study), Denmark, FPS (Finnish Public Sector study), NWCS (Netherlands Working Condition Survey) and POLS (Permanent Onderzoek Leefsituatie/Continuous Survey on Living Conditions), the Netherlands, the Whitehall II study (WHII), UK, and the WOLF-Stockholm study, Sweden.

Participants with complete data on ERI scales were included in these analyses, yielding an analytic sample of 31,790 employed men and women. The individual questionnaire items of the partial proxy scales available from the other 10 studies are presented in *Table 1*.

Statistical analyses

We applied the following procedure. First, we restricted the analysis to the two scales ‘effort’ and ‘reward’, measuring

the extrinsic components of the ERI model. Here, the original short versions were used from the 5 studies with 3 items measuring ‘effort’ and 7 items measuring ‘reward’ (referred to as ‘complete scales’). In accordance with the core theoretical notion, we additionally constructed a ratio of the two scales to quantify the degree of mismatch between effort and reward at individual level, according to established procedures (Siegrist et al. 2004, 2009; Leineweber et al. 2010; Li et al. 2012a). By dichotomizing the effort-reward ratio (ER-R) at the cut-point of 1.0 (where efforts and rewards are equal), an exposure measure of work stress was available for all individuals.

Proxy measures of the two scales were then defined from available information of the 10 remaining studies. Given a relatively high degree of heterogeneity of the wording of items and their response format, different scales were defined by consultation of the investigators who were responsible for the studies. As a result, 2 partial ‘effort’ scales, 7 partial ‘reward’ scales, and 7 partial ER-R measures were identified (*Table 1*). Items and scales are available on request. The specific partial proxy scales available in each IPD cohort study are as follows: Belstress (Effort = A/Reward = E), DWECS and COPSOQ-2 (B/A), FPS (B/B), IPAW (B/D), NWCS and POLS (B/G), PUMA and WOLF-Stockholm (B/C), WHII (B/F). Finally, the partial proxy scales were replicated in the five studies available for validation, and full and partial scales were compared.

Mean scale values were computed for each participant and for all complete and partial scales. Associations between ‘effort’ and ‘reward’ scales were ascertained using Pearson’s correlation coefficients with accompanying 95 % confidence intervals using Fisher’s transformation. Sensitivity, specificity, and Kappa (κ) statistics were calculated to evaluate the agreement between the different ER-R measurements. The following interpretations of the Kappa statistics were given: 0.00–0.20 indicates slight agreement, 0.21–0.40 fair, 0.41–0.60 moderate, 0.61–0.80 good/substantial agreement, and 0.81–1.00 a very good/almost perfect agreement (Landis and Koch 1977).

At last, we examined associations between the alternative ERI scales and health as a relevant theoretical criterion for scale validity. A single-item measure of self-rated health (SRH) was available for all studies. Using logistic regression, we analysed if high stress in terms of ER-R values >1 of the respective original or proxy scale was related to poor self-rated health (defined as very bad, bad, neither good nor bad health). Regressions were conducted separately for each of the five studies, and random-effects meta-analysis summary estimates were calculated over these results.

All analyses were performed using SAS version 9.2 (SAS Institute Inc., Cary, NC, USA), except meta-analyses we conducted with Stata/MP, version 11.1.

Table 1 Items in the complete and partial effort/reward scales

ERI Questionnaire		Partial scales ^a													
		Complete scale ^b	Effort A	Effort B	Reward A	Reward B	Reward C	Reward D	Reward E	Reward F	Reward G				
<i>Effort</i>															
ERI 1	I have constant time pressure due to a heavy work load.	X	X	X											
ERI 2	I have many interruptions and disturbances while performing my job	X	X												
ERI 6	Over the past few years, my job has become more and more demanding	X	X												
<i>Reward</i>															
ERI 7	I receive the respect I deserve from my superiors	X			X									X	
ERI 11	My job promotion prospects are poor	X			X									X	X
ERI 12	I have experienced or I expect to experience an undesirable change in my work situation	X			X					X					
ERI 13	My employment security is poor	X			X					X				X	
ERI 15	Considering all my efforts and achievements, I receive the respect and prestige I deserve at work	X			X				X					X	
ERI 16	Considering all my efforts and achievements, my job promotion prospects are adequate	X												X	
ERI 17	Considering all my efforts and achievements, my salary/income is adequate	X			X				X				X	X	X

^a Partial scales from BELSTRESS, Belgium, COPSQ-2 (Copenhagen Psychosocial Questionnaire Study), DWECs (Danish Work Environment Cohort Study), IPAW (Intervention Project on Absence and Well-being Study) and PUMA (Copenhagen Burnout Inventory Study), Denmark, FPS (Finnish Public Sector study), NWCS (Netherlands Working Condition Survey) and POLS (Permanent Onderzoek Leefsituatie/Continuous Survey on Living Conditions), the Netherlands, the Whitehall II study (WHII), the UK, and the WOLF-Stockholm study, Sweden

^b This is the standard short form of the effort/reward imbalance questionnaire. The long version additionally includes the following items: ERI 3: 'I have a lot of responsibility in my job', ERI 4: 'I am often pressured to work overtime', ERI 5: 'My job is physically demanding', ERI 8: 'I receive the respect I deserve from my colleagues', ERI 9: 'I experience adequate support in difficult situations', ERI 10: 'I am treated unfairly at work', ERI 14: 'My current occupational position adequately reflects my education and training'

Results

Sample characteristics

Concerning the socio-demographic composition, the 5 study samples were comparable with respect to their mean age. In 3 studies, approximately half of the participants were women, whereas their proportion was lower in the GAZEL and WOLF-F studies (Table 2).

Correlation between partial and complete scales

Pearson's correlation coefficients between the complete scales 'effort' and 'reward' and the alternative partial proxy versions of these scales are displayed in Table 3. In all 5 studies, the two partial proxy scales measuring effort showed high correlation ($r > 0.92$), thus representing the original measurement equally well. More variation was observed across the 7 partial reward scales. Scales A and B demonstrate a high degree of accordance between all 5 studies, with all coefficients ranging between $r = 0.90$ and $r = 0.98$. Again, these partial scales containing 4 or 5 of the standard items represent the original scale to a highly satisfactory degree. To a somewhat lesser extent, this is the case for the partial scales E and F where coefficients range from $r = 0.83$ to $r = 0.93$. These scales contain only 3 out of 7 standard items of the original scale. The relationships between the complete scale and the partial proxy scales C, D, and G were weaker, but still acceptable with coefficients ranging from 0.72 to 0.87. It should be noted that the reward scale G is restricted to 2 standard items, thus representing only part of the more comprehensive original construct. The strength of these associations was reasonably similar across the 5 studies. Taken together, there is statistical evidence to support the use of the 2 partial scales of 'effort' and the 7 partial scales of 'reward' as proxies of the respective original scales.

Table 2 Characteristics of study participants ($N = 31,790$)

Study	N^a	Mean age (SD)	Female, N (%)
HNR	1,838	53.4 (5.0)	749 (40.8)
SOEP	10,387	43.2 (10.7)	4,696 (45.2)
SLOSH	5,048	47.3 (10.8)	2,690 (53.3)
GAZEL	9,713	51.9 (3.0)	2,568 (26.4)
WOLF-F	4,804	45.6 (10.2)	1,348 (28.1)

HNR Heinz Nixdorf Recall Study, Germany; SOEP The German Socio-Economic Panel Study; SLOSH Swedish Longitudinal Occupational Survey of Health, Sweden; GAZEL the GAZ and Electricité Study, France; WOLF-F Work, Lipids, and Fibrinogen, Norrland Follow-up study, Sweden

^a Participants with no missing data on ERI ratio

Specificity and sensitivity

Table 4 shows the results of analysis of sensitivity and specificity for ER-R based on partial proxy scales using complete scales as the criterion. Sensitivity and specificity ranged between 59–93 and 85–99 % depending on the scale and study. High agreement is observed for ER-R based on effort scale B and reward scales A and F. To a lesser extent, this is the case for ER-R from effort scale B combined with reward scales B, D, or G. Agreement for effort scale B—reward scale C and effort scale A—reward scale E is lower, but still acceptable, in particular concerning the sensitivity analyses. In all analyses, specificity measures are substantially higher than sensitivity measures. Kappa statistics varied for the different scales with good agreement in four of the seven alternative ratios. For the combination of the effort B scale with the reward scales C, D, and G, agreement was generally good, with single kappa values indicating only moderate agreement.

Associations with self-rated health

In accordance with the theoretical assumptions, the original 'full' effort-reward ratio is strongly associated with poor self-rated health (Fig. 1). The results for five of the seven proxy scores are very close to the result for the original ER-R. Odds ratios for the ER-R scores calculated with the reward proxies F and G are slightly lower but still significantly elevated. Study specific analyses (not shown) revealed a high degree of agreement between the results from the single studies.

Discussion

This study is the first to assess psychometric properties of abbreviated versions of the scales 'effort' and 'reward' of the effort-reward imbalance model of work stress, drawn from several large European cohort studies. We observed very high correlations between the complete and proxy effort and reward scales if at least 2 effort and 4 or 5 reward items were included. Correlations were still high if at least 3 reward items were measured, but lower in one version (G) where only 2 reward items were available. Moreover, in a cross-sectional analysis, the complete scale and all partial scales were strongly associated with poor self-rated health. Overall, the partial scales show acceptable degrees of agreement and may be used as proxy measures of the original scales in future analyses of pooled data from all 15 available studies.

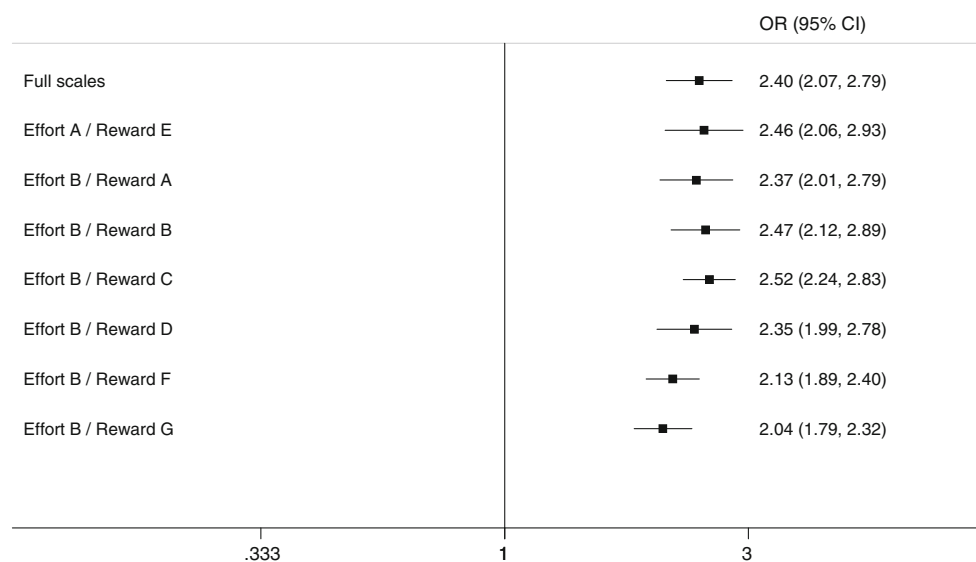
We also calculated the sensitivity, specificity, and kappa coefficients of the different effort-reward ratios derived from partial scales by comparing them with the ratios of the

Table 3 Correlation coefficient (*r*) between the complete effort and reward scales versus the alternative partial versions of these scales

Version of partial scale	Correlation with complete scale (Pearson's <i>r</i> , 95 % confidence interval) by study				
	HNR	SOEP	SLOSH	GAZEL	WOLF-F
Effort					
Scale A	0.927 (0.920–0.933)	0.937 (0.935–0.939)	0.949 (0.946–0.951)	0.946 (0.943–0.948)	0.920 (0.915–0.924)
Scale B	0.927 (0.920–0.933)	0.933 (0.930–0.935)	0.950 (0.948–0.953)	0.944 (0.942–0.946)	0.929 (0.925–0.932)
Reward					
Scale A	0.953 (0.949–0.957)	0.965 (0.963–0.966)	0.976 (0.975–0.978)	0.970 (0.969–0.972)	0.949 (0.946–0.952)
Scale B	0.904 (0.895–0.912)	0.932 (0.929–0.934)	0.941 (0.938–0.944)	0.930 (0.928–0.933)	0.903 (0.898–0.908)
Scale C	0.809 (0.792–0.824)	0.851 (0.846–0.856)	0.863 (0.856–0.870)	0.825 (0.819–0.831)	0.759 (0.747–0.771)
Scale D	0.801 (0.784–0.816)	0.848 (0.842–0.853)	0.871 (0.864–0.878)	0.832 (0.826–0.838)	0.797 (0.786–0.807)
Scale E	0.869 (0.857–0.880)	0.908 (0.904–0.911)	0.926 (0.922–0.930)	0.923 (0.920–0.926)	0.886 (0.880–0.892)
Scale F	0.830 (0.815–0.843)	0.887 (0.882–0.891)	0.903 (0.898–0.908)	0.932 (0.930–0.935)	0.872 (0.865–0.879)
Scale G	0.720 (0.697–0.741)	0.799 (0.792–0.806)	0.840 (0.831–0.848)	0.857 (0.851–0.862)	0.778 (0.766–0.790)

Table 4 Sensitivity, specificity, and kappa values of the dichotomised effort/reward imbalance ratio (>1 vs. ≤1) based on partial scales versus complete scales by study

Partial scales	Sensitivity/specificity/kappa compared to complete scales by study				
	HNR	SOEP	SLOSH	GAZEL	WOLF-F
Effort A—Reward E	63.4/98.5/0.68	69.8/98.2/0.74	69.7/98.8/0.76	69.8/98.9/0.76	69.1/97.2/0.72
Effort B—Reward A	90.1/95.0/0.69	89.6/95.6/0.80	88.7/97.1/0.86	85.1/97.9/0.84	85.4/94.5/0.78
Effort B—Reward B	83.8/95.8/0.69	81.3/96.2/0.76	81.8/97.6/0.82	70.1/98.9/0.77	79.7/95.9/0.77
Effort B—Reward C	71.1/94.2/0.55	73.4/96.3/0.71	67.5/98.1/0.72	64.9/98.3/0.71	58.9/97.2/0.63
Effort B—Reward D	76.1/93.5/0.56	76.8/94.6/0.68	77.5/96.0/0.75	64.4/98.0/0.70	73.7/94.2/0.69
Effort B—Reward F	82.4/94.2/0.62	88.6/92.5/0.71	93.2/90.1/0.74	90.3/94.1/0.77	91.5/88.7/0.71
Effort B—Reward G	73.2/94.6/0.58	84.4/90.7/0.64	91.1/88.5/0.70	84.9/92.6/0.71	89.5/84.6/0.64

**Fig. 1** Summary estimates for the associations between effort-reward imbalance and poor self-rated health for different ERI ratio scales (odds ratios and 95 % confidence interval)

complete scales. This is important in view of the contribution of this combined measure towards explaining health risks. The ratio quantifies the imbalance between effort and reward at individual level and, thus, is expected to contribute to the explanation of adverse health effects above and beyond the effects attributed to the individual scales (Siegrist et al. 2004; Leineweber et al. 2010; Li et al. 2012a, b; Magnavita et al. 2012). Again, taking the diverse quality of the proposed partial scales into account, the findings of this analysis demonstrated acceptable degrees of concordance across the different combinations of scales (Table 4).

A major strength of the study is the inclusion of data from a large population providing information on effort-reward imbalance at work from more than 30,000 employed men and women recruited in the frame of 5 epidemiological studies from different European countries. Given the high number of observations, a high degree of statistical precision is expected to result from respective comparative analyses. While the findings of this study offer a strong basis of generalisation beyond those derived from single cohort studies, one should nevertheless keep in mind that the samples were recruited from Northern, North-western, or Western European countries. It is therefore not known to what extent they hold true in other parts of Europe and in other continents. A further limitation relates to potential bias produced by differences in the meaning of wording of single items in the translation of original scales. Although established quality criteria are usually applied in these translation processes, potential differences in meaning may reduce a strict comparability of results, in particular in case where scales were reduced to 2 or 3 items. A further limitation specifically concerning the shorter scales is that not all of the theoretical dimensions of the ERI model are captured by respective items (e.g. all dimensions of reward). This shortcoming could reduce the content validity of the respective scale. Moreover, the precision of comparison was reduced by differences in the wording and in the response format of some items from proxy measures.

There are only a few previous studies which compared ERI scales across different countries. Most notable among those is a study funded by the European Science Foundation which included data from France, Belgium, Sweden, Germany, and the United Kingdom (Siegrist et al. 2004). As a majority of these studies used the complete scales, analytical emphasis was put on the factorial structure of the scales, testing goodness of fit by confirmatory factor analysis, as well as on content validity, estimating odds ratios of poor self-rated health according to the effort-reward ratio. Again, a high degree of comparability of results was obtained. Another study compared the original ERI scales and a proxy measure within a Swedish working population, again with

reasonable comparability (Fahlen et al. 2004). Moreover, in-depth analyses based on item response theory were conducted by Tsutsumi et al. (2009) using cross-national data from different studies. The results indicated a sufficient cross-country comparability of standard scales.

In conclusion, in view of prevailing constraints in applying extensive, time-consuming original scales measuring established work stress models in large-scale epidemiological studies, our findings support the notion that short proxy measures of the original scales measuring the effort-reward imbalance model can be used for data analysis in work stress research. Clearly, this is always a second best choice, but the trade-off between strong comparability of exposure measures on the one hand and robustness and precision of prediction of health outcomes on the other hand needs to be considered.

Acknowledgments The authors would like to thank Dirk De Bacquer, Annalisa Casini, Els Clays, France Kittel (Belstress), Jakob B Bjorner, Marianne Borritz, Hermann Burr, Ida E H Madsen, Martin Lindhardt Nielsen, Jan H Pejtersen (COPSOQ-I, COPSOQ-II, DWECS, IPAW, PUMA studies), Jaana Pentti, Paula Salo, Tuula Oksanen, Jussi Vahtera, Marianna Virtanen (FPS), Goedele A. Geuskens, Wendela E. Hooftman, Irene I.L.D. Houtman (POLS, NWCS), Peter JM Westerholm (WOLF S), G. David Batty, Jane E. Ferrie and Michael G. Marmot (Whitehall II) for IPD-Work Consortium collaboration and providing information on effort-reward items available in the respective studies. The IPD-Work consortium is supported by the EU New OSH ERA research programme (funded by the Finnish Work Environment Fund, Finland, the Swedish Research Council for Working Life and Social Research, Sweden, the German Social Accident Insurance, Germany, the Danish National Research Centre for the Working Environment, Denmark), the Academy of Finland (grant #132944), the BUPA Foundation (grant 22094477). MK is supported by the Medical Research Council, UK, and the US National Institutes of Health (R01HL036310; R01AG034454), and the Academy of Finland.

Ethical approval Each constituent study in the IPD-Work consortium was approved by the relevant local or national ethics committees, and all participants gave informed consent to take part. Details of the ethical approval in each study are provided in Appendix 1.

Conflict of interest The authors declare that they have no conflict of interest.

References

- Belkic KL, Landsbergis PA, Schnall PL, Baker D (2004) Is job strain a major source of cardiovascular disease risk? *Scand J Work Environ Health* 30:85–128
- Blekesaune M, Solem PE (2005) Working conditions and early retirement: a prospective study of retirement behavior. *Res Aging* 27:3–30
- Cartwright S, Cooper CL (eds) (2009) *The Oxford handbook of organizational well-being*. Oxford University Press, Oxford
- Dragano N (2007) *Arbeit, stress und krankheitsbedingte Frührenten: Zusammenhänge aus theoretischer und empirischer Sicht. [Work, stress and health-related early retirement: theoretical and empirical analyses]* VS Verlag

- Elovainio M, Kivimäki M, Vahtera J (2002) Organizational justice: evidence of a new psychosocial predictor of health. *Am J Public Health* 92:105–108
- Fahlen G, Peter R, Knutsson A (2004) The Effort-Reward Imbalance model of psychosocial stress at the workplace—a comparison of ERI exposure assessment using two estimation methods. *Work Stress* 18:81–88
- Fransson EI, Nyberg ST, Heikkilä K, Alfredsson L, De Bacquer D, Batty GD et al (2012) Comparison of alternative versions of the job demand-control scales in 17 European cohort studies: the IPD-Work consortium. *BMC Public Health* 12:62
- Greenberg J (2010) Organizational injustice as an occupational health risk. *Acad Manag Ann* 4:205–243
- Head J, Kivimäki M, Siegrist J, Ferrie JE, Vahtera J, Shipley MJ, Marmot MG (2007) Effort-reward imbalance and relational injustice at work predict sickness absence: the Whitehall II study. *J Psychosom Res* 63:433–440
- Karasek R, Theorell T (1990) *Healthy work. Stress, productivity, and the reconstruction of working life*. Basic Books, New York
- Kuper H, Singh-Manoux A, Siegrist J, Marmot M (2002) When reciprocity fails: effort-reward imbalance in relation to coronary heart disease and health functioning within the Whitehall II study. *Occup Environ Med* 59:777–784
- Landis JR, Koch GG (1977) The measurement of observer agreement for categorical data. *Biometrics* 33:159–174
- Leineweber C, Wege N, Westerlund H, Theorell T, Wahrendorf M, Siegrist J (2010) How valid is a short measure of effort-reward imbalance at work? A replication study from Sweden. *Occup Environ Med* 67:526–531
- Li J, Loerbroks A, Jarczok MN, Schöllgen I, Bosch JA, Mauss D et al (2012) Psychometric properties and differential explanation of a short measure of effort-reward imbalance at work: A study of industrial workers in Germany. *Am J Industrial Med*. doi:10.1002/ajim.22018
- Li J, Loerbroks A, Shang L, Wege N, Wahrendorf M, Siegrist J (2012b) Validation of a short measure of effort-reward imbalance in the workplace: evidence from China. *J Occup Health*. doi:10.1539/joh.12-0106-BR
- Magnavita N, Garbarino S, Siegrist J (2012) The use of parsimonious questionnaires in occupational health surveillance: psychometric properties of the short Italian version of the effort/reward imbalance questionnaire. *ScientificWorldJournal*. doi:10.1100/2012/372852
- Nieuwenhuijsen K, Bruinvels D, Frings-Dresen M (2010) Psychosocial work environment and stress-related disorders, a systematic review. *Occup Med* 60:277–286
- Nyberg ST, Heikkilä K, Fransson EI, Alfredsson L, de Bacquer D, Bjorner JB et al (2012) Job strain in relation to body mass index: pooled analysis of 160,000 adults from 13 cohort studies. *J Intern Med* 272:65–73
- Salavecz G, Chandola T, Pikhart H, Dragano N, Siegrist J, Jockel KH et al (2010) Work stress and health in Western European and post-communist countries: an East-West comparison study. *J Epidemiol Community Health* 64:57–62
- Schnall PL, Dobson M, Roskam E (2009) *Unhealthy work: causes, consequences, cures*. Baywood Pub. Co., Amityville
- Siegrist J (1996) Adverse health effects of high effort/low reward conditions. *J Occup Health Psychol* 1:27–41
- Siegrist J, Starke D, Chandola T, Godin I, Marmot M, Niedhammer I, Peter R (2004) The measurement of effort-reward imbalance at work: European comparisons. *Soc Sci Med* 58:1483–1499
- Siegrist J, Wege N, Pühlhofer F, Wahrendorf M (2009) A short generic measure of work stress in the era of globalization: effort-reward imbalance. *Int Arch Occup Environ Health* 82:1005–1013
- Stansfeld S, Candy B (2006) Psychosocial work environment and mental health—a meta-analytic review. *Scand J Work Environ Health* 32:443–462
- Stephens A, Kivimäki M (2012) Stress and cardiovascular disease. *Nat Rev Cardiol*. doi:10.1038/ncardio.2012.45
- Tsutsumi A, Kawakami N (2004) A review of empirical studies on the model of effort-reward imbalance at work: reducing occupational stress by implementing a new theory. *Soc Sci Med* 59:2335–2359
- Tsutsumi A, Iwata N, Watanabe N, de Jonge J, Pikhart H, Fernández-López JA, Xu L, Peter R, Knutsson A, Niedhammer I, Kawakami N, Siegrist J (2009) Application of item response theory to achieve cross-cultural comparability of occupational stress measurement. *Int J Methods Psychiatr Res* 18:58–67