Methods for the Systematic Review on Worklessness and Wellbeing Concluded in September 2016

The systematic review that forms the basis for this meta-analysis sought studies that focused on any form of worklessness (not being in regular employment or education/training, because of unemployment, retirement, ill-health or disability, or family care), exits from work and wellbeing. Before performing the review, the research team developed a protocol that outlined the structure of the intended systematic review such as the research questions proposed, and the criteria set for inclusion of the relevant studies. This protocol was registered on PROSPERO (reference number: WITHELD TO PRESERVE ANONYMITY)[1] and followed the best practice PRISMA-P reporting guidelines (Shamseer et al., 2015).

Inclusion and Exclusion Criteria

The review team employed the "Population, Intervention, Comparators, Outcome, Study design" [PICOS] approach (Liberati et al., 2009; Shamseer et al., 2015) when developing the research questions and setting criteria for including or excluding studies for the review. The strategy employed could be described as follows:

Population: we were interested in research about well-being and the working-age population in general, including the recently retired. Our review sought research undertaken in a developed economic context, for example EU-15 countries, USA or Australia.

Intervention: We were mainly interested in studies that investigated the relationship between worklessness and wellbeing. Therefore, the state of worklessness was our intervention of interest.

---

[1] The protocol is publicly available and can be found in http://www.crd.york.ac.uk/PROSPERO using the reference number.

Comparators: The review aimed to identify a range of factors which might have influenced the relationship between wellbeing and worklessness. Accordingly, studies with any or no comparators were included. Studies that focused or provided comparisons across the different wellbeing outcomes of different worklessness states were eligible for inclusion. Moreover, we sought studies that shed light on varying effects of worklessness across different sub-groups (e.g. comparisons based on age-groups; gender, those with disabilities or health conditions, family status, and region).

*Outcomes***:** Acknowledging the potential reverse causality, this review focused on any sort of change in wellbeing as a result of worklessness, its duration and the associated transitions. Therefore, the outcome that the review was interested in was changes in wellbeing. We focused on the effects of worklessness on various indicators of wellbeing such as life satisfaction, mental health, stress, anxiety and depression**.**

Study Designs: We sought evidence from empirical research – quantitative, qualitative or mixed methods studies. Studies that use longitudinal methods were preferable. However, we also sought evidence from high quality cross-sectional studies that were based on econometric techniques that are known to tackle the endogeneity issues (for example, regression discontinuity, difference in difference, instrumental variable approach).

*Other***:** The studies were not excluded based on their publication type, however, we only included peer-reviewed empirical research. This was because of the volume of data available within the peer-reviewed research, which ensured sufficient high-quality evidence was available to answer the questions set out in our review. We were interested in studies that involved an empirical analysis. Therefore, review articles or policy reports were excluded from our review. Although non-English papers were not excluded, the search was performed in English language data bases. Initially, there was no date restriction. However, given the large number of search results, we excluded studies published earlier than 1990. We believed that

this would still allow us to answer our research questions based on a sufficiently long period of time, which included recessions affecting different types of workers.

Search Strategy and Electronic Databases

The search terms were developed according to the research questions of the review and inclusion/exclusion criteria explained above. In order to see the validity of the search terms, dummy searches were performed initially and, advice was sought from relevant experts in the field. The final set of search terms were as follows:

Keywords for Intervention (worklessness): *(return_to_work* OR get_ready_to_work* OR return_to_employ* OR retire* OR unemploy* OR disab* OR workless* OR progress*_into_employment OR progress*_into_work OR transition*_into_work OR transition*_out_of_work OR transition*_into_employ* OR transition*_out_of_employ* OR return_to_labour OR return_to_labor OR economically_inactive OR duration_of_unemploy*)*

AND

Keywords for Outcome (well-being): *(stress OR well-being OR wellbeing OR well_being OR emotion* OR affect* OR mood OR job_satis* OR life_satis* OR worthwhile* OR eudaimon* OR happ* OR anxiety OR depress* OR self-esteem OR self-efficacy OR mental_health OR mental_ill* OR psychological_health OR resilience)*

AND

Keywords for Study Design/Methods: *(difference*_in_difference* OR regression_discontinuity OR matching_methods OR instrumental_variable* OR propensity_score_matching OR longitudinal OR panel OR pseudo-panel OR quasi-experiment OR experiment OR randomised_control_trial OR randomized_control_trial OR interview OR case study OR qualitative)*

In order to identify the existing systematic reviews and meta-analyses that investigated the relationship between worklessness, exits from work and wellbeing, the research team performed an initial scoping review using the same search terms for the intervention and outcome keywords, but the study design/methods keywords were replaced by (best_evidence_review* OR systematic_review* OR meta-analys*).

The search was performed using the Web of Science, PsycInfo, Scopus, PubMed, Academic Search Complete, Business Source Complete and Econlit.

Study Selection

The search based on the databases described above yielded 4,028 studies. Exclusion of all the duplicates, non-peer reviewed published work and studies published before 1990 resulted in 1,961 titles – hits – that needed to be sifted according to the agreed exclusion/inclusion criteria. Sifting was performed independently by two review authors. Any disagreements between the two reviewers were discussed and, if consensus was not reached, the study was included for the next stage. Reviewers always chose to put through the study in question to the next stage even when there was only a weak chance of meeting the inclusion criteria for the sake of not missing any information that might be relevant for the review. The strength of agreement between the reviewers were very good with Cohen's Kappa of 0.876 (SE of kappa = 0.012 95% confidence interval: From 0.852 to 0.899).

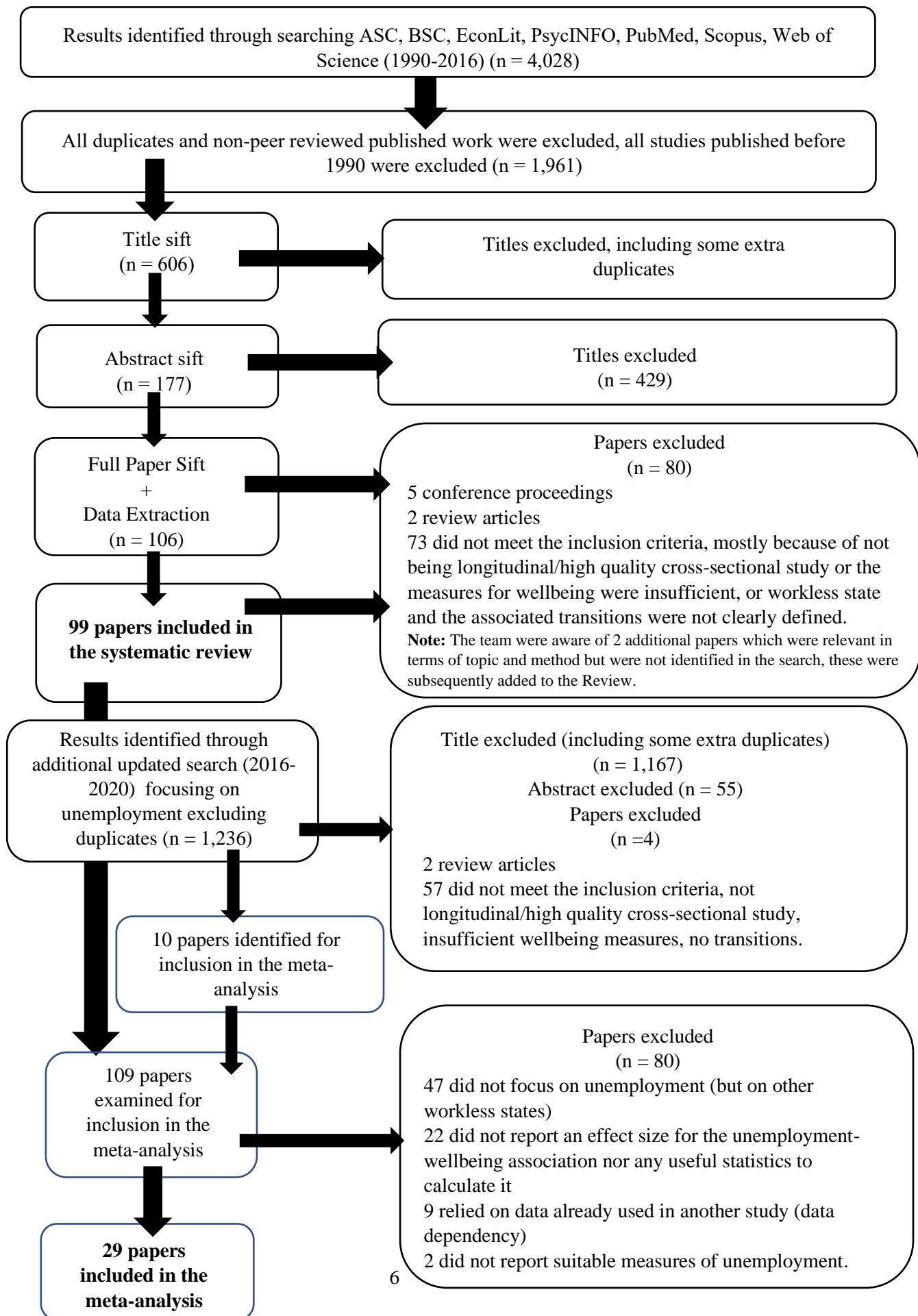In the next stage, abstracts were sifted according to the inclusion criteria. Any study that two reviewers were in doubt about including/excluding were double-checked by a third member of the team. Studies that were selected during the abstract sift stage were then evaluated as full papers and assessed by the reviewers, Figure 1 summarises the selection of the papers and the rationale for any exclusions.

Updating the Evidence Base: Additional Search Concluded in November 2020

In accordance with the aim of our meta-analysis, as explained in the paper, we selected studies from the review outlined above that considered "unemployment" as a workless state. Given that this review was concluded in 2016, we updated the evidence base with an additional search which we concluded in November 2020 following the same systematic review methods detailed above. This update of the evidence base produced an additional 1,236 titles covering studies published between 2016 to November 2020. Following the same steps explained above, review authors selected 10 studies which met our selection criteria and focused on unemployment. Six of these studies were retained for meta-analysis.

Further details for inclusion/exclusion from this update of the search are presented in Figure 1.

Figure 1. Stages of study selection

Results identified through searching ASC, BSC, EconLit, PsycINFO, PubMed, Scopus, Web of Science (1990-2016) (n = 4,028)

All duplicates and non-peer reviewed published work were excluded, all studies published before 1990 were excluded (n = 1,961)

Title sift
(n = 606)

Titles excluded, including some extra duplicates

Abstract sift
(n = 177)

Titles excluded
(n = 429)

Full Paper Sift
+
Data Extraction
(n = 106)

Papers excluded
(n = 80)

5 conference proceedings
2 review articles
73 did not meet the inclusion criteria, mostly because of not being longitudinal/high quality cross-sectional study or the measures for wellbeing were insufficient, or workless state and the associated transitions were not clearly defined.
Note: The team were aware of 2 additional papers which were relevant in terms of topic and method but were not identified in the search, these were subsequently added to the Review.

99 papers included in the systematic review

Results identified through additional updated search (2016-2020) focusing on unemployment excluding duplicates (n = 1,236)

Title excluded (including some extra duplicates)
(n = 1,167)
Abstract excluded (n = 55)
Papers excluded
(n = 4)

2 review articles
57 did not meet the inclusion criteria, not longitudinal/high quality cross-sectional study, insufficient wellbeing measures, no transitions.

10 papers identified for inclusion in the meta-analysis

109 papers examined for inclusion in the meta-analysis

Papers excluded
(n = 80)
47 did not focus on unemployment (but on other workless states)
22 did not report an effect size for the unemployment-wellbeing association nor any useful statistics to calculate it
9 relied on data already used in another study (data dependency)
2 did not report suitable measures of unemployment.

29 papers included in the meta-analysis

6

Supplementary Material B

The Coding Process

For each primary study, we coded any statistical information on the association between unemployment and wellbeing that could be used to compute Cohen's d. Hence, we extracted correlations, mean differences, means, standard deviations and sample sizes for the employed and unemployed groups, odd ratios, regression coefficients (standardized, unstandardized, type of model), 2x2 contingency tables, *chi* square (with $df = 1$) and total sample size, and any further statistics used. Additionally, in order to correct for measurement error in the dependent variable (i.e., wellbeing), as explained below, we coded the reliabilities of the various wellbeing measures (i.e., the Cronbach's alpha coefficients of the wellbeing variables) for each primary study, when reported. For those papers including more than one measure of wellbeing, statistics were coded separately for each of them.

Inter-rater reliabilities (r_{IR}) were calculated for the statistics used to compute effect sizes, specifically for correlations ($r_{IR} = 1$), odds ratios ($r_{IR} = 1$), standardized regression coefficients ($r_{IR} = .99$), means, standard deviations and sample sizes of the unemployed ($r_{IR} = .99$, $r_{IR} = .99$, $r_{IR} = .55$, respectively) and employed groups ($r_{IR} = .99$, $r_{IR} = .99$, $r_{IR} = .81$, respectively). Discrepancies, often resulting from typos, were resolved through discussion. Inter-rater reliabilities were also computed for the continuous moderator variables (i.e., age, level of education, publication year, first year of data collection, and length of the dataset), and were all equal to 1, except for number of measurement occasions ($r_{IR} = .84$). Inter-rater agreement (Cohen's kappa) for the categorical moderator country was .67. Again, disagreements were solved via discussion.

Details of Correction for Unreliability in the Wellbeing Measure

To correct for unreliability in wellbeing, when a primary study contained information (i.e., Cronbach's alpha) about the reliability of the wellbeing measures (e.g., life satisfaction, depression, anxiety), we used these estimates. Additionally, with regard to life satisfaction,

we employed the reliability coefficients from the study by Lucas and Donnellas (2012) for the BHPS and GSOEP datasets, by Wanous, Reichers, and Hudy (1997) for the ECHP dataset, and by Butterworth and Crosier (2004) for the HILDA dataset. For one primary study using the Center for Epidemiological Study Depression Scale (CES-D, Radloff, 1977), we retrieved the reliability coefficient for the scale (Howe et al., 2012) and tailored it to the length of the questionnaire of the primary study (i.e., 15 items) *via* the Spearman-Brown formula (Schmidt & Hunter, 2015). Six studies used the General Health Questionnaire (GHQ-12; Goldberg, 1972) as a measure of psychological wellbeing, but only four of them reported reliability coefficients. Thus, a bare-bone meta-analysis was run, yielding a meta-analytic reliability estimate of .90 ($K = 4$, $N = 161,445$) (Schmidt & Hunter, 2015).

Measures of Depression other than GHQ-12 Used in the Primary Studies

CES-D, Radloff, 1977; Composite International Diagnostic Interview – Short Form, CIDI-36, Kessler et al., 1998; Diagnostic Interview Schedule for Depression, DIS, American Psychiatric Association, 1987; Mini International Neuropsychiatric Interview (MINI), Sheehan et al., 1998; Short-Form Health Survey, SF-36, Ware et al., 2000; Symptom Checklist-Core Depression scale, SCL-CD$_6$, Magnusson Hanson et al., 2009.

Publication Bias Checks

To test for the possible effect of publication bias, we applied the funnel plot technique (Sterne & Egger, 2005) to each of the principal five general meta-analyses based on 10 studies or more (i.e., M1, M2, and M3 in Table 1), as recommended by Kepes et al. (2012). This plots the precision of the effect size (the inverse of the standard error) of each primary study on the y-axis as a function of these effect sizes on the X-axis, and it examines the distribution for asymmetry, which denotes the presence of bias. A symmetrical distribution was observed for the meta-analysis focusing on life satisfaction as the wellbeing outcome (M2, Table 1). Some asymmetry was exhibited for the other two meta-analyses. To assess the extent of the bias and the potential influence on the conclusions, we employed the trim and

fill method (Duval & Tweedie, 2000) to impute the number of missing samples needed to correct for asymmetry and to estimate the meta-analytical effect size if those samples were not missing. Furthermore, we implemented Egger's test of the intercept (Egger et al., 1997), where the bias is verified by using precision to predict the standardized effects. In the regression equation, the bias is captured by the significance of the intercept, which should pass through the origin in the case of a symmetric distribution. Specifically, the trim and fill technique suggested that 8 studies were missing in the meta-analysis including all the wellbeing measures (M1 in Table 1) and 2 studies were necessary in the meta-analysis centred on mental health outcomes (M3 in Table 1) to reach symmetry. However, the reduction in the effect sizes calculated through trim and fill was negligible, since the new estimates fell into the 95% confidence interval of the original meta-analyses. Moreover, the Egger's intercept test did not find a significant intercept for any of the two meta-analyses ($p = .31$; $p = .49$ respectively).

For the sake of precision, the publication bias analyses were also performed separately for the moderator subgroups based on a minimum of 10 samples (i.e., M6, M10, M11, M12, M14, M15, M16, M17, and M18 in Table 2), as asymmetry can be caused by moderating effects present in the analyses (Kepes et al., 2012). In the meta-analyses centred on long-term unemployment (M18, Table 2) and including those countries characterised by high OECD employment protection index (M17, Table 2) and low GII (M10, Table 2) levels, the samples were symmetrically distributed, and no missing studies were imputed through the trim and fill technique. Asymmetry was reduced for the remaining subgroup meta-analyses. Specifically, the trim and fill procedure imputed 3 samples necessary to reach symmetry in the subgroup meta-analysis centred on women only (M6, Table 2); 2 and 4 samples for the two meta-analyses distinguishing between non-European and European studies, respectively (M14 and 15, Table 2); 4 samples for the subgroup meta-analysis on countries with low

OECD employment protection index (M16, Table 2); and 4 and 3 samples for the subgroup meta-analyses including studies characterised by low NTW (M12, Table 2) or high GII (M11, Table 2) scores, respectively. However, in all cases the imputed effect sizes fell into the 95% confidence interval of the original meta-analyses. Finally, Egger's intercept test did not yield significant in any of the six subgroups (M6: $p = 09$; M14: $p = .27$; M15: $p = .36$; M16: $p = .44$; M12: $p = 30$; M11: $p = 23$). In sum, we did not find strong evidence for publication bias.

References

Butterworth, P., & Crosier, T. (2004). The validity of the SF-36 in an Australian National Household Survey: Demonstrating the applicability of the Household Income and Labour Dynamics in Australia (HILDA) Survey to examination of health inequalities'. *BMC Public Health*, *4*(1), 44. https://doi.org/10.1186/1471-2458-4-44.

Duval, S. J., & Tweedie, R. L. (2000). Trim and fill: A simple funnel plot–based method of testing and adjusting for publication bias in meta-analysis. *Biometrics, 56*(2), 455-463. https://doi.org/10.1111/j.0006-341x.2000.00455.x.

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal, 315*, 629-634. https://doi.org/10.1136/bmj.315.7109.629.

Goldberg, D. (1972). *The detection of psychiatric illness by questionnaire*. London: Oxford University Press.

Howe, G.W., Hornberger, A. P., Weihs, K., Moreno, F., & Neiderhiser, J. M. (2012). Higher-order structure in the trajectories of depression and anxiety following sudden involuntary unemployment. *Journal of Abnormal Psychology*, *121*(2), 325–338. https://doi.org/10.1037/a0026243.

Kepes, S., Banks, G. C., McDaniel, M. & Whetzel, D. L. (2012). Publication bias in the

organizational sciences. *Organizational Research Methods, 15*, 624–662.

https://doi.org/10.1177/1094428112452760.

Kessler, R. C., Andrews, G., Mroczek, D. K., Ustun, B., & Wittchen, H.U. (1998). The

World Health Organization composite international diagnostic interview short-form

(CIDI-SF). *International Journal of Methods in Psychiatric Research*, *7*(4), 171–185.

https://doi.org/10.1002/mpr.47.

Liberati A., Altman D.G., Tetzlaff J., Mulrow C., Gøtzsche P.C., Ioannidis J.P.A., et al.

(2009). The PRISMA Statement for Reporting Systematic Reviews and Meta-

Analyses of Studies That Evaluate Health Care Interventions: Explanation and

Elaboration. *PLoS Med, 6*(7), e1000100.

https://doi.org/10.1371/journal.pmed.1000100.

Lucas, R. E., & Donnellas, M. B. (2012). Estimating the reliability of single-item life

satisfaction measures: Results from four national panel studies. *Social Indicators

Research*, *105*, 323–331. https://doi.org/10.1007/s11205-011-9783-z.

Magnusson Hanson, L. L., Theorell, T., Bech, P., Rugulies, R., Burr, H., Hyde, M.,

Oxenstierna, G., & Westerlund, H. (2009). Psychosocial working conditions and

depressive symptoms among Swedish employees. *International Archives of

Occupational and Environmental Health*, *82*(8), 951-960.

https://doi.org/10.1007/s00420-009-0406-9.

Radloff, L. S. (1977). The CES-D Scale: A self-report depression scale for research in the

general population. *Applied Psychological Measurement*, *1*(3), 385-401.

https://doi.org/10.1177/014662167700100306.

Schmidt, F. L., & Hunter, J. E. (2015). *Methods of meta-analysis. Correcting error and bias

in research findings*. (3rd ed.). Thousand Oaks, CA: Sage.

Shamseer, L., Moher, D., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., Shekelle, P., &
Stewart, L.A. (2015). Preferred reporting items for systematic review and meta-
analysis protocols (PRISMA-P) 2015: Elaboration and explanation. *British Medical
Journal*, *349*, g7647. https://doi.org/10.1136/bmj.g7647.

Sterne, J. A., & Egger, M. (2005). Regression methods to detect publication bias and other
bias in meta-analysis. In H. R. Rothstein, A. J. Sutton, & M. Borenstein (Eds.),
Publication bias in meta-analysis: Prevention, assessment, and adjustments (pp. 99-110).
Chichester, UK: Wiley.

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

Ware, J. E., Snow, K. K., & Kosinski, M., 2000. *SF-36 Health Survey: Manual and
Interpretation Guide*. Lincoln, RI: Quality Metric Incorporated.

Table SM1

Results of the correlational moderator analyses (for the continuous moderators)

Moderator	Corrected correlation	p	K
Age	-0.35	0.41	13
University degree (% sample w/)	-0.33	0.50	11
Higher education (% sample w/)	0.25	0.53	15
Lower education (% sample w/)	0.69	0.14	11
HDI	-0.50	0.13	21
GII	0.44	0.17	22
NTW	-0.22	0.49	23
Publication year	-0.11	0.72	27
Length of dataset	-0.16	0.58	26
First year in dataset	-0.20	0.50	26

Note. HDI = Human Development Index; GII = Gender Inequality Index; NTW = Norm to Work.