UNIVERSITY OF OXFORD

Wellbeing Research Centre

Working Paper Series

2305

Estimating effects of individual-level workplace mental wellbeing interventions: Cross-sectional evidence from the UK

William Fleming

April 2023

Estimating effects of individual-level workplace mental wellbeing interventions: Cross-sectional evidence from the UK

William Fleming*
william.fleming@hmc.ox.ac.uk
Wellbeing Research Centre, University of Oxford

April 2023

Abstract

Promotional mental health interventions have become increasingly common in British workplaces but evaluative research lacks methodological quality, adequate sample size, or is more concerned with economic outcomes than employee health. Recently, a critical scholarship has emerged in sociology that questions the labour and health politics of workplace 'wellness'. This article presents quantitative research addressing these concerns by estimating the 'treatment effect' of mental health programmes using clustered Bayesian propensity score analysis with the 'Britain's Healthiest Workplace' survey – a sample of 143 British organisations and 27,932 workers. The analysis estimates the effect of a range of common initiatives including: mindfulness, resilience training, stress management and wellbeing apps. No evidence is found to demonstrate that these strategies improve worker mental health across multiple employee mental health measures. This suggests that recent critical literature is right to be concerned and that convenient well-being strategies should be given less financial and institutional support.

Keywords: Human Resources, Interventions, Management, Mental health, Mindfulness, Resilience, Wellbeing

*This research was funded by the ESRC and RAND Europe (Award reference: ES/J500033/1). Data was provided by VitalityHealth. I would like to thank Adam Coutts, Brendan Burchell, Chris van Stolk, Diane Coyle, Maria Iacovou and Tarani Chandola for feedback on this research. Thanks to Isabelle Stuart for editorial assistance.

1 Introduction

The promotion of wellbeing is commonplace in the contemporary British workplace. The best estimates for the prevalence of corporate action on workers' wellbeing comes from the Chartered Institute of Personnel and Development (CIPD, 2022) who survey employers annually. Their most recent data has over half of UK employers reporting a 'formal wellbeing strategy', with another third providing at least ad hoc support. The most common practices include employee assistance programmes (EAPs), counselling, and health promotion, such as advice on healthy lifestyles. These counts have steadily rose over the past decade, and it now seems safe to assume that a majority of British workers have some exposure to health and wellbeing narratives and policies.

It was the publication of Black's (2008) independent review for the UK Department for Work and Pensions, *Working for a healthier tomorrow*, that is said to have 'galvanised' the now widespread concern for workers' wellbeing (Kowalski & Loretto, 2017: 2230). The report has three key objectives: preventing illness through the promotion of health and wellbeing; early workplace intervention on health problems; and, where possible, supporting those out of work to improve labour market participation. Black's role as National Director for Health and Work represents how work and health had converged in public policy agendas.

Yet this policy focus has narrowed in more recent years, emphasising mental health as the central concern. This has only been amplified by the Covid-19 pandemic – a catalyst for public and corporate awareness of wellbeing. To stick with UK government reports and policy, Stevenson & Farmer's (2017) *Thriving at Work* review represents this shift towards mental health. Their 'vision' includes core standards of creating good quality jobs; empowering individuals to manage their own mental health; equipping employers to promote good mental health and manage bad mental health in the workforce; and a reduction in unemployment. The official response was clear: 'all employees' mental health should be taken care of in the workplace' (DWP & DHSC, 2017: 14).

Both Stevenson & Farmer and Black's recommendations therefore urge employers to promote wellbeing, so what does this consist of practically? I have already mentioned EAPs, health promotion, and counselling, but the latest guidelines on mental wellbeing offer broader advice for employers (National Institute for Care Excellence – NICE, 2022). They recommend that employers 'adopt a tiered approach', including organisation-level, individual-level, and targeted approaches. Organisational responses change the structure of how work structured, aiming to minimise sources of stress and insecurity; whereas individual-level responses seek change in workers' capacities, attitudes, and behaviours. While NICE explicitly state 'do not use individual-level approaches to replace organisational strategies', they do advocate universal access to mindfulness, yoga, and meditation programmes. It is these individual-level approaches that are most common in business advice, practice and evidence on how to promote workplace mental wellbeing.

The benefits of individualistic approaches to promoting wellbeing have been extensively researched, as well as increasingly debated. There is a large scholarship of experimental work testing the effects of participation in programmes like stress management, mindfulness, health promotion, and more. These have been collected and published in multiple systematic reviews, including from NICE (2022). Despite the apparent scale of evidence examining the question of effectiveness, many technical issues remain, such as small samples, non-generalisable populations, and a lack of engagement with selection bias and attrition.

Beyond specific questions of research design and reporting, the field is also plagued by poor quality research. This is best captured by the research conducted by consultants at Deloitte in support of Stevenson & Farmer's review (Hampson et al., 2017). They collate economic evaluations of individual-level initiatives like regular mental health screening and lifestyle advice, declaring an 'overwhelmingly positive' business case for investment in workforce wellbeing. However, their ad hoc selection process and inability to identify valid studies undermines return-on-investment estimates that are widely and uncritically shared in public media and academic literature. Evidence such as this must be interrogated.

These points of contention form the beginnings of an empirical critique of individual-level mental wellbeing interventions, suggesting that there is insufficient evidence in support of the wide-spread adoption and promotion of such practices. However, alongside empirical concerns is a strategic and sociological critique, with many criticisms levelled at such individualistic practices' propagating, in the words of the Trades Union Congress (TUC, 2013, 2018), 'changing the worker, and not the workplace'. Looking also to sociological literature, Frayne's (2019) concerns of 'bad prescriptions, faulty diagnoses, and toxic side-effects' argue a lack of effectiveness, a misidentification of problems, and potential unintended harm. Other critiques go further, with some commentators sharing the view that mental health and wellbeing initiatives are more interested in social control than with improving workers' wellbeing (e.g. Davies, 2015; Foster, 2018; Murphy & MacMahon, 2022).

With this policy context, as well as these empirical and normative critiques in mind, this article will contribute to the evidence base for workplace mental wellbeing interventions. The interventions I evaluate here are individual-level, promotional rather than ameliorative, and rely on active participation. They are strategies that offer individualised, behavioural support for mental health, with examples like mindfulness, resilience and stress management training, and time management training. Results are derived from observational data collected through the 'Britain's Healthiest Workplace' survey. Using this sample, I conducted clustered propensity score analysis to estimate cross-sectional 'treatment effects' on participating workers' subjective wellbeing. The straightforward research question asks what is the average treatment effect on the treated for individual-level workplace mental wellbeing interventions? Next, by exploring organisation variance and group-level effects for gender, ethnicity, income, and prior stress levels, I include expanded analysis on the question of 'what works for whom and in which circumstances?' (Nielsen & Miraglia, 2017). As a point of comparison for the individual-level interventions, I also share predicted relationships between key dimensions of job quality and workers' wellbeing.

An observational study like this is unavoidably limited in several ways for causal inference. Results are correlational, but, when interpreting the results, I provide a comprehensive discussion of the effects that considers selection bias from reverse causation, i.e. participation in interventions may be prompted by prior levels of wellbeing. While these findings can never act as definitive causal evidence, the BHW data nevertheless offers an invaluable opportunity to estimate possible treatment effects across multiple organisations, with multiple interventions, and where fidelity is not as assured as in experimental trials (Fikretoglu et al., 2022). I position these results as supplementary to necessary experimental research. Additionally, I make an original effort to consider sociological accounts in the evaluation of the statistical results.

2 Literature review

2.1 Existing evidence base

In Stevenson & Farmer's (2017: 6) review, they declare that 'there is a pressing need for more evidence', and more recently, The Wellcome Trust, launching research funding for mental wellbeing at work, state that there are 'significant gaps in the evidence base' with 'limited causal evidence' for interventions (Newman, 2021). One problem is that the collection of evidence has been of poor quality. As an example, we can take Deloitte's review (Hampson et al., 2017) in support of Stevenson & Farmer, noted in the introduction. While not an academic piece of work, its importance for policy demands it receive critical attention.

The review evaluates the return-on-investment (ROI) for mental health interventions, estimating returns of 4:1 (Hampson et al., 2017). This report is presented as 'systematic review', but follows no established methodology of systematic review. They rely on a narrow focus on ROI estimates, inadequate search terms, and poorly defined selection and exclusion criteria. From their process, seven 'primary sources' were selected, from which the average ROI was determined. These primary studies, their selection, and Deloitte's evaluation suffer from a series of technical problems.

First, there is circular citations and duplicated results. Matrix (2013) and Knapp et al. (2011) are both included despite both only being reviews that repeat the same results from Mills et al. (2007). The Mills study is a controlled trial evaluating a single firm's health promotion programme, an initiative which centred on providing workers with information on healthy lifestyles. A second 2011 Knapp et al. study is cited as another primary source, but I am unable to locate that report and no references are provided. Also included as a primary source for the ROI estimates is Hargrave & Hiatt's (2005) study evaluating an EAP for depression, cited with the incorrect publication date, despite it using financial costs from another study, not the programme they were investigating. In Deloitte's review, there are two other primary sources: a simulation of hypothetical interventions which again does not use original data (PWC, 2014), and an evaluation of depression tele-therapy on a US health insurance plan (Wang et al., 2007). This systematic review then contains two studies with original results for wellbeing interventions, but which evaluate interventions with very different levels of intensity – wellness information and counselling for depression. This is the summation of the ROI estimates which were declared 'overwhelmingly positive' by its authors and by (Stevenson & Farmer, 2017: 27), and which are regularly cited uncritically in media and academic literature.

There are only two further pieces of academic evidence that are cited in the Stevenson & Farmer review. The first is a meta-analysis of existing systematic reviews (Wagner et al., 2016) which also suffers from peculiar screening decisions. They report four reviews that find positive evidence for mental wellbeing interventions. These include: a review of OCD and PTSD treatments (Noordik et al., 2010); a review of studies on employee access to mental health therapy (Pomaki et al., 2012); and, despite mental health being the target outcome, reviews of neck pain (Aas et al., 2011), and lower back pain (Tveito et al., 2004) interventions. The inclusion of these inappropriate studies in Wagner et al.'s review is unexpected, and undermines their overall positive conclusions. The final source Stevenson & Farmer provide is an RCT of a mental health training course for managers in the Australian fire and rescue service (Milligan-Saville et al., 2017). A relevant intervention, but its results on key outcomes are not statistically significant, they have a small treatment sample (n = 44), and its study population consists of workers in extrinsically dangerous, strenuous, and traumatic jobs.

It is no surprise that Stevenson & Farmer suggest there is a need for more evidence considering the substandard selection of evidence conducted in support of their report. The systematic reviews suffer from 'garbage in, garbage out', and the additional RCT that is cited indicates issues in sample size and external validity. While the cited evidence used in support of this UK government report is inadequate, there is more rigorous analysis of individual-level strategies elsewhere.

In the evaluation literature there is debate about the effectiveness of individual-level strategies. In a narrative review, Tetrick & Winslow (2015) argue that they are beneficial for both preventative and ameliorative approaches to stress, but that higher quality research is still required. Yet systematic reviews typically have cautioned against the benefits of individual-level approaches. Daniels et al. (2021b) suggest that there can be positive effects, but results are strongly dependent on the context, implementation, and mechanisms. LaMontagne et al. (2007) are more critical in their review, stating that while individual-level strategies can be effective, they often are not, and that the effectiveness 'disappears over time'. They argue organisational level approaches have benefits for the individual and the organisation, offering a preventative strategy for stress especially.

Specific types of individual-level interventions have also been covered by systematic reviews, with resilience training programmes one such example. Robertson et al. (2015) claim that resilience training 'can improve personal resilience and is a useful means of developing mental health and subjective wellbeing in employees'. Despite these positive claims, all included studies have small samples, several of which are less than 50 workers across treatment and control groups (e.g. N = 16 in Burton et al. (2010); N = 29 in Pipe et al. (2012); N = 12 in Sherlock-Storey et al. (2013); N = 32 in Sood et al. (2011). Further, several studies again specifically target workers with extrinsically stressful and traumatic jobs, such as the police, or armed forces. While small samples and specific populations do not necessarily undermine positive results, they do challenge the external validity.

Another popular intervention is mindfulness, which has its own evidence base with several corresponding systematic reviews. Bartlett et al. (2019); Joyce et al. (2018) & Vonderlin et al. (2020) all conclude that there are positive effects for subjective wellbeing, and NICE (2022) similarly find enough evidence to recommend universal access for all British workers. However, the primary limitation in mindfulness and resilience intervention studies that consistently goes unmentioned is treatment selection bias. As an example, we can take Bostock et al.'s (2019) widely-cited evaluation of a digital mindfulness app in two UK organisations. They find positive effects for a range of subjective wellbeing measures and biomarkers. However, participants in the treatment were those who replied to an email advertising this initiative, with participants randomised after this recruitment phase, and with no interrogation of any potential bias. The danger is that these strategies may prove beneficial for those who are believers, but that many do not see them as worthwhile. This would explain the participation rates of approximately 13% of workers across both sites. With results reporting average treatment effect on the treated, the organisational impact of the intervention is therefore an order of magnitude smaller.

Concerns for treatment selection bias concur with recent organisation-wide studies of wellness programmes. Jones et al.'s (2019) RCT of a multi-component wellness programme identified few discernible benefits for participating workers, finding that the people who did use and engage with practices were those who were already healthy. Song & Baicker's (2019) similar analysis, this time of a multi-site strategy, found no benefits for wellbeing either. Both Jones et al.'s and Song & Baicker's studies have caused controversy. The quality of research design is high for both, so it is significant that they find null effects for wellbeing strategies that the researchers themselves devised with best practice and high fidelity in mind. Song & Baicker are co-authors on a previous study showing high ROI for promotional strategies (Baicker et al., 2010) which is extremely highly cited. Both sets of authors acknowledge that their findings sit in opposition to much of the literature, which shows overwhelming positive effects; consequently, they raise questions about the quality of that existing literature, and both have faced criticism from prominent advocates of behavioural strategies (e.g. Goetzel, 2020). The stories around these studies suggest that higher quality research is needed, and that there may be existing vested interests at play.

Strategies like mindfulness and resilience training are also investigated outside of workplaces. Reviewers of these interventions are positive about the potential of mindfulness (e.g. Creswell, 2017). While these results indicate benefits of these practices as adjunctive treatment for depression (Eisendrath et al., 2016) and anxiety (Hoge et al., 2013), there is less support for these interventions as preventative strategies. Where there is positive evidence, sample size is a concern, as is the overestimation of effects (Goldberg et al., 2021). University settings are one context where they are effective for student wellbeing (Galante et al., 2020; Hood et al., 2021), but effects are small and treatment selection bias is again evident. These results are relevant, but it is doubtful how comparable these more general results are for the context of the workplace. Paid employment is a complex social dynamic of consent, coercion, and reliance, with the work environment playing a fundamental role in shaping workers' health and wellbeing.

2.2 Sociological critiques

I have touched on the criticism of individual-level interventions from empirical researchers that argue they do not take into account workplace environment, and may ultimately blame individuals for their stress. For these reasons, research teams such as Daniels et al. (2021b); Fox et al. (2022) & Lovejoy et al. (2021) argue that individual-level interventions are unlikely to be effective. Veld & Alfes (2017) also suggest that HRM researchers too often take 'the optimistic approach' when it comes to wellbeing practices, failing to fully appreciate the harm that could be caused.

The empirical critique is echoed by critical accounts of workplace wellbeing. For example, Hull & Pasquale (2018) argue that wellness programmes are ineffective for improving health and reducing costs, but they cherry-pick their own evidence. Frayne (2019: 12) develops the criticism in more detail, arguing that individual-level strategies entail 'faulty diagnoses, bad prescriptions and toxic side effects'. Faulty diagnoses are caused by ignoring the social context of health and wellbeing, especially the physical and social environment of work. Accusations of bad prescriptions highlights an insinuation that the problem is inside individual workers, rather than caused by external factors. Both are said to produce the toxic side effects of self-blame and 'phony empowerment' and undermine the intentions of interventions, causing decline in mental wellbeing. Reflecting on Frayne's comments, the danger is in interventions treating workers' wellbeing as though it is isolated from working conditions, as well as failing to appreciate the normative and political dimensions of wellbeing.

The political dimensions of wellbeing interventions include the strategic aims of such practices. Firstly, trade unions worry that wellbeing practices are the latest attempt to circumvent the responsibilities of unions (Foster, 2018; TUC, 2013). There are concerns from researchers that 'health and wellbeing' obscures 'health and safety' (Sorensen et al., 2021). Questions are also raised over whether wellbeing practices merely offer a 'reputational alibi' that absolve organisations of responsibility, allowing employers to claim they are making efforts to improve workers' wellbeing, regardless of actual outcomes (Southwood, 2019). This criticism would place wellbeing interventions within a cynical understanding of the optics of corporate social responsibility.

Most critical sociological accounts emphasise a normative critique of these strategies, linked closely to the idea of 'bad prescriptions' that seek change in the individual. Many commentators worry that workplace mental wellbeing interventions are an attempt to construct an 'ideal' worker by intervening in the subjectivity of workers, endeavouring to make them more engaged, more productive, and more amenable to corporate goals and managerial demands (Davies, 2015; Foster, 2018; Frayne, 2019; Murphy & MacMahon, 2022). In doing so, health and wellbeing are seen as merely instrumental for profit in a new model of social control.

To some extent, qualitative interviews support the sociological critique, especially when it comes to health promotion. Holmqvist & Maravelias's (2011) analysis of a Swedish firm suggests that the incorporation of workplace health promotion as a component of organisational control can have both positive and negative effects. Zoller's (2004) findings from interviews with workers report similar divergence between workers and management, and (Wallace, 2019) develops a theory of 'productive sickness' from his qualitative findings. More closely linked with the mental health, practitioners of mindfulness at work acknowledge that they moderate ambiguity in the meaning of mindfulness to match dominant managerial perspectives (Islam et al., 2022). Doing so turns mindfulness into an 'empty signifier', i.e. a linguistic encoding whereby opposition in meaning is deliberately obscured along power and class dynamics.

Most of these critiques of workplace wellbeing interventions suggest that such practices are not effective. However, the claim that practices are managerial techniques of social control does not necessarily discount potential improvements in subjective wellbeing measures. I will refer back to these sociological accounts in my later discussion on the results.

3 Data and Methods

3.1 Data

For this analysis I used the 2018 wave of the Britain's Healthiest Workplace (BHW) survey. The BHW is a multi-level repeated cross-sectional survey, with data collected at both employee and organisation level. At the employee level, individual workers provide information on their lifestyles, health behaviours, and physical and mental health. For the organisation level, a senior manager or HR representative responds to represent the entire organisation. Questionnaires cover the general characteristics of the organisation, their internal strategies for promoting wellbeing, and the organisation's perception performance and wellbeing. The repeated survey waves were conducted between 2013 and 2019. There is a small panel sample, but it was not made available by the commercial data owners, VitalityHealth. The total sample size of the 2018 wave is 27,932 individual employees clustered in 153 workplaces. I reduced the sample to 27,919 employees in 143 UK workplaces, excluding organisations with less than five respondents as they would not provide reliable estimates.

The BHW provides a convenience sample with selection bias at both levels, and is therefore not representative of the UK national workforce or any single organisation. Participating organisations are not randomly sampled and must opt in, while I also assume that those organisations that do opt in are those with an existing corporate interest in wellbeing. Financial and insurance services are slightly over-represented in the survey coverage. Respondents are also all voluntary at the employee level, and I make the similar assumption that those who complete the survey are those most engaged in wellbeing discourses and practices. Internal response rates vary by organisation size, with larger employers having lower response rates on average. At the individual level, women, younger workers, those on mid-to-high incomes, and white workers are all over-represented.

Despite these sample limitations, no other large-scale individual-level survey of the UK's workplace wellbeing landscape exists, meaning the BHW provides a unique opportunity to quantitatively analyse the effects of interventions across multiple organisations. Further, the target population for this analysis is narrower than the entire British workforce, instead focussing on those workers who participate in promotional workplace wellbeing programmes. For the same reason, I applied no additional weighting.

The BHW has been used in two peer-reviewed studies before. (Daniels et al., 2021a) use data from the 2014 and 2015 waves of the BHW to inspect any differences in reported wellbeing scores among those who participated in a range of health and wellbeing programmes. They do find positive effects, but participation did not buffer against psychosocial work hazards. Their study is undermined by including a range of interventions on the same scale when evaluating treatment effects. These grouped interventions vary in intensity between simple health information up to individual therapy for mental illness. Fida et al.'s (2021) study suffers from the same limitation, with different types of intervention grouped together. Both Daniels et al. and Fida et al.'s studies also use the BHW sample but without any weighting, matching or modelling of selection bias.

3.1.1 Interventions

Table 1 presents the wellbeing interventions evaluated in this study, including participation rates for each. Following a qualitative process of typology building for interventions, I selected these interventions from a longer list included in the BHW survey. All the interventions are promotional activities that target the individual, and which require active participation. These programmes are all popular initiatives in British workplaces (CIPD, 2022).

When analysing the dimensions of job quality, I recoded a series of variables in the BHW to resemble binary treatments for comparing with the wellbeing interventions. All question wording and recoding are included in Table A.2, with most originally consisting of 5-point Likert scales.

3.1.2 Participation

Participation in the wellbeing interventions was measured by employee self-reports. Survey respondents are faced with lists of common wellbeing initiatives and are required to answer yes/no to three related questions: whether their employer offers a programme, whether they have participated in the previous twelve months, and whether they believe it had a positive effect. Table A1 shows that participation in all programmes appears low, with only around a quarter of the sample participating in any of the possible interventions.

3.1.3 Predictors of participation

To achieve balance in the sample through propensity scores, I included many predictors of treatment assignment. Individual-level variables were gender, age, ethnicity, job level, carer status, marital status, existing mental health condition, contract, work shifts, working hours, work environment, commute, tenure, participation in wellbeing monitoring, feeling respected, whether health is important for company success, and whether a supervisor is concerned with wellbeing. Organisation-level predictors were union recognition, organisation size, industry, and whether incentives for participation are offered. β coefficients for the estimated effect on participation for each variable is included in Figure A.1. I do not include income at the balancing stage because of missing values, but I do include it as an interaction in the later analysis. An extra predictor of prior wellbeing was used to model treatment selection bias: 'during the last 12 months have you felt unwell as a result of work-related stress?', with answers of 'no', 'yes, to some extent', and 'yes, definitely'.

Table 1: Availability and participation rates for mental wellbeing interventions

Intervention	Not available Count (%)	No Count (%)	Yes Count (%)	%* No:Yes
Any mental health promotion programme	**11,447 (41)**	**9,653 (34.6)**	**6,819 (24.4)**	**58.6:41.4**
Volunteering or charity work	18,394 (65.9)	6,393 (22.9)	3,160 (11.2)	66.9:30.1
Mindfulness classes or programmes	21,807 (78)	4,524 (16.2)	1,617 (5.8)	73.7:26.3
Resilience, energy or stress management classes or programmes	23,121 (82.8)	3,650 (13.1)	1,178 (4.1)	75.6:24.4
Well-being app for physical health, mental health and lifestyle issues	24,442 (87.5)	2,381 (8.5)	1,124 (4)	67.9:32.1
Massage or relaxation classes or programmes	23,439 (83.9)	3,447 (12.3)	1,063 (3.8)	76.4:23.6
Workload or time management training	23,142 (82.9)	4,153 (14.9)	653 (2.2)	86.4:13.6
Financial well-being courses or programmes	24,071 (86.2)	3,456 (12.4)	423 (1.4)	89.1:10.9
Events promoting healthy sleep	26,792 (95.9)	857 (3)	299 (1.1)	74.1:25.9
Apps/programmes promoting healthy sleep	26,817 (96)	837 (3)	294 (1)	74:26
Coaching (one-on-one sessions on mental health and wellbeing)	25,686 (92)	1,976 (7)	288 (1)	87.3:12.7
Online coaching	26,943 (96.5)	863 (3.1)	142 (0.5)	85.9:14.1

Note: Programmes are sorted from high to low for participation. ∗ Excluding 'not available'.

3.1.4 Outcomes

I adopted a multiple outcome approach for understanding the effect at the individual level. The Short Warwick-Edinburgh Mental Well-Being Scale (SWEMWBS) was the most appropriate measure of wellbeing, and I present results using this scale as the primary outcome. The SWEMWBS is a construct of self-reported measures for subjective wellbeing. The SWEMWBS is a well-established psychometric scale for evaluating policies with the scale emphasising the positive dimensions of mental health, making it appropriate for the promotional strategies evaluated here. Higher values correspond to higher wellbeing. When interpreting the results, the SWEMWBS is standardised where I report β coefficients, but the full 35-point scale is used for predicted values.

I also tested several additional measures: Kessler scores,[1] life satisfaction (0-10), job satisfaction (0-10), the work engagement,[2] and self-rated mental health (5-point Likert). For comparability I standardised all measures. My selection of subjective measures was designed to reflect the effect on workers themselves, taking these interventions on the terms they are presented, rather than through an exploration of indirect relationships.

3.2 Analytic strategy

To determine the effect of the interventions on workers' wellbeing, I conducted clustered Bayesian propensity score analysis (PSA). Traditional PSA was designed to infer a causal 'treatment effect' when handling cross-sectional observational data (Rosenbaum & Rubin, 1983; Rubin, 1997). PSA provides a reduced sample that allows a cross-sectional sample to better resemble a randomised experiment, modelling treatment selection bias to an extent. PSA with matching operates in two stages: design and analysis. For the design stage, predictors are selected on the basis that they will theoretically affect treatment assignment and the outcome. A binomial logistic regression model is fitted to estimate the propensity score, i.e. a scalar of the conditional probability of treatment assignment. The score is then used in a process of matching, where cases are paired to provide a sample with balanced treatment and control groups. In the analysis stage, the so-called treatment effects are estimated by comparing treatment groups.

Adopting a Bayesian estimation approach is practically productive because of its applicability for estimating hierarchical models with variable group sizes, because it can better manage multiple levels of uncertainty (Western, 1999; Zhang et al., 2016). Bayesian PSA is still in a developmental stage with ongoing epistemic and empirical debates meaning a procedure is not established.[3] I used Kaplan & Chen's (2012) two-step Bayesian PSA ('BPSA-2'). BPSA-2 uses Bayesian MCMC estimation at the design stage for propensity scores, it then matches the sample using a preferred matching method, before again performing MCMC

[1]The Kessler Psychological Distress Score measures negative aspects of mental health through ten questions. Higher scores correspond to poorer wellbeing but are inverted for comparison.

[2]Utrecht Work Engagement Scale-9 (Schaufeli & Bakker, 2004)

[3]An optimal Bayesian matching method would match using the curve or distribution tiles of propensity score predictions from the MCMC chain. Such an approach has not yet been developed in methodology literature.

estimation for treatment effects. Using MCMC estimation means all coefficients are posterior means with 95% highest density credible intervals. MCMC chain lengths were 25,000 iterations for the design stage and 5,000 iterations for the analysis stage following consultation of the Raftery-Lewis diagnostics. Prior beliefs in treatment effects of the interventions were uncertain from an inconclusive literature review and, therefore, I specified priors using frequentist likelihood estimation with a chain burn-in length of 500. For random slope models, I manually set non-informative priors to ensure model convergence.

To pair individuals with another in their organisation, I augmented the process using within-cluster matching.[4] Ideally, there would be an understanding of different selection biases in different clusters, meaning a combination of within and across cluster matching would be possible, but this is not possible with the BHW and within-cluster matching is superior to ignoring firm grouping. Figure A.2 shows the balance of the sample before and after matching, while Table A.1 includes the post-matching sample.

There are three methodological problems with PSA and matching methods generally that are relevant here: unobservables, sample balance, and causal direction. First, matching can only be conducted on observed variables, therefore likely missing important contextual features. I included a comprehensive range of confounders in the design model, but this limitation likely persists. The second problem of sample balance is particular to PSA. King & Nielsen (2019) argue that propensity scores should not be used for matching because in trying to resemble a randomised trial, the sample ultimately ends up unbalanced. However, in this analysis the post-matching sample provided excellent balance across the predictors. As a robustness check, I searched the balance of the post-matching sample for all interventions and predictors, finding negligible, non-significant differences. The third problem with PSA is the remaining potential for reverse causality due to the cross-sectional data. This is the major limitation of this analysis, with wellbeing outcomes likely predicting the participation in interventions resulting in remaining treatment selection bias. For example, a worker or a team who are over-stressed may be given access to a mindfulness course for support. To model this selection bias, I used interaction terms with the self-reported measure of whether respondents suffered from work-related stress in the last year to estimate whether their participation in the intervention was driven by their prior wellbeing. While this extended modelling does provide some additional confidence, the potential for reverse causation remains, seriously undermining all causal claims. I discuss all results with consideration of the remaining bias.

In summary, I present results from clustered Bayesian PSA. I used random intercept binomial logistic regression models to calculate the propensity score for each intervention. I then matched on the propensity score within organisations. The matched sample was then used to predict the average coefficient for a series of outcome variables. Random intercepts were used to incorporate organisation-level variance, with results from this model presented in Figure 1. I also estimated random slopes models to explore differences between organisations (Table A.3).

For additional results, I provide grouped treatment effects estimated through interaction effects for gender, ethnicity, and income. I selected gender and ethnicity because I found

[4]'CMatching' R package (Cannas, 2019).

both to have sociologically interesting differences in participation rates, as well as both proving relevant for sociological understandings of how wellbeing and organisations may be gendered and racialised. Income is crucial for understanding experience of work. For this stage, I grouped interventions by effect direction with models for positive, negative and null effects. I estimated different models for each interaction term and for each directional set. Interaction effects were also estimated for whether an individual suffered from work-related stress. I present interactions as predicted SWEMWBS scores.

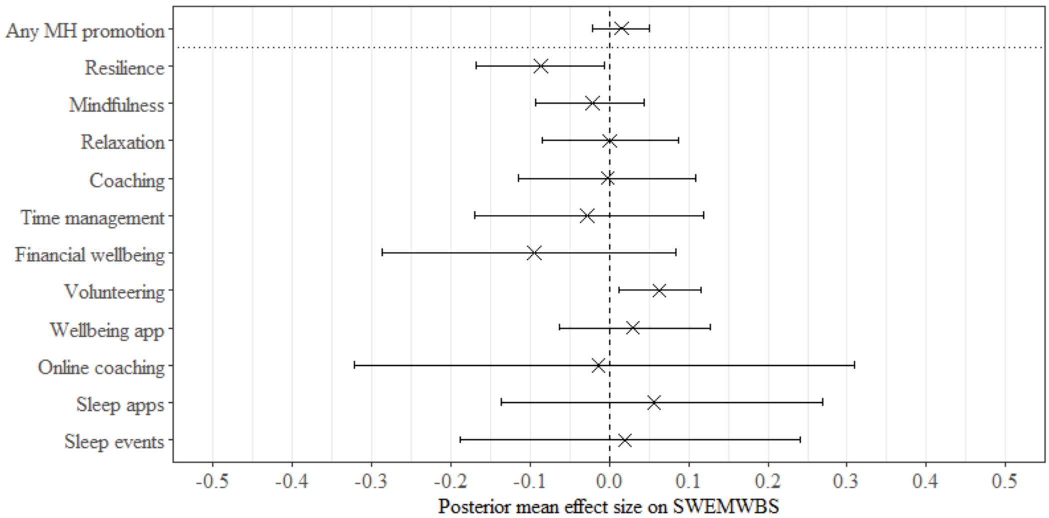
Finally, I also evaluated the relationships for a series of dimensions of job quality and wellbeing. These results are intended as a point of comparison for the interventions. I present coefficients for the unmatched and unweighted sample, as well as for propensity score matched samples. The former is a more coherent approach where there is not a clearly discernible intervention, but for comparing results, I also include matched-sample analysis.

4 Results

4.1 Interventions

Figure 1 shows the estimated 'treatment effects' for the mental wellbeing interventions. The primary finding is that there is no reliable effect on mental wellbeing. Results presented use a standardised SWEMBS as the outcome variable, but results are similar for the other measures of mental wellbeing (Figures A.3, A.4, A.5, A.6 & A.7). The effect was estimated for all interventions combined (top point in Figure 1) to represent the average effect for the

Figure 1: 'Treatment effect' of workplace mental wellbeing interventions on workers' SWEMWBS



Note: SWEMWBS is standardised ($\bar{y}$ = 0, std.dev. = 1). β Coefficient estimates are posterior means from MCMC chains with length 5,000. Whiskers show 95% highest density credible intervals.

average intervention.

For the following specific types of interventions, estimates indicate null effects on workers' mental wellbeing: mindfulness, massage and relaxation classes, time management, coaching, financial wellbeing programmes, wellbeing apps, online coaching, sleep apps, and sleep events. Volunteering is the only type of intervention to suggest positive effects on workers' wellbeing. I estimate a negative association between wellbeing and resilience and stress management programmes.

Larger credible intervals are estimated where there are smaller sample sizes, indicating more volatility in these results. Bayesian analysis with the BHW convenience sample cannot provide power estimations, but the following discussion and conclusions will focus on those interventions with larger numbers of participants (N $>$ 500, Table A.1), where credible intervals are small enough to identify plausible directional effects.

For variance at the organisational level, I also estimated random slopes for participation. There is negligible variance in these random effects, indicating that effects are similar across organisations (Table A3). The largest organisation-level variance is for resilience ($\sigma^2_{u1} =$ 0.029; CIs: 0.007, 0.076) and time management ($\sigma^2_{u1} = 0.022$; CIs: 0.004, 0.076), highlighting how low these estimates are. These results indicate that the null effects were consistent across organisations, and that random effects were not an accurate model for the data. This lack of between-organisation difference reiterates the null effects at the individual level. Results imply that it is not the case that effects are averaged out across contexts, but that there are no effects across the sample.

SWEMWBS is the primary outcome but the additional results are in the Appendix. Job satisfaction (Figure A.5) and life satisfaction (Figure A.6) show similar results to SWEMWBS, with null effects for overall participation and many of the interventions. For these outcomes, the relationship for resilience is again negative, but the estimate for mindfulness programmes is negative as well. Again, volunteering has a positive coefficient. For Kessler scores (Figure A.3), the picture looks similar, but the 'any intervention' estimate is negative. However, the specific interventions that have negative coefficients are coaching, time management, and, again, resilience and stress management.

For subjective work engagement (Figure A.4), the results appear somewhat different. All interventions combined are estimated to have a positive association, but this appears to be linked with the positive effects for volunteering opportunities and, less so, wellbeing apps and therefore can be considered consistent with the other results. For further work engagement estimates, a negative estimate is found for mindfulness interventions (β ¡ 0.01), but not resilience. The negative estimates for mindfulness indicate similar mechanisms as the resilience interventions as these types of intervention regularly overlap.

4.2 Gender, ethnicity, income and prior work-stress interactions

Interaction terms offer insight for possible group-level treatment effects. Figures 2 3 & 4 show the predicted wellbeing scores for the interaction term models for the 'null effect interventions', resilience interventions, and volunteering interventions, respectively.

Examining gender and ethnicity, the null effect interventions again show no notable differences between participants and non-participants. Predictions for income levels do suggest possible differences: for high-earners (greater than £40,000) who do participate, predicted wellbeing is lower than non-participants; whereas for lower earners, I again found no relationship.

Results are similar for both resilience and volunteering interventions, with no notable group-level differences in the predictions for gender and ethnicity. Among higher earners who do participate in resilience programmes, there is some suggestion of a similar relationship to the null effects interventions, where participants have slightly lower predicted wellbeing than non-participants. However, intervals overlap and differences appear only small.

To model treatment selection bias caused by prior levels of wellbeing, I included whether an individual suffered from work-related stress as an interaction. For the null effect interventions (Figure 2), the predicted values again show no differences within stress groups. For those who answer 'no' to suffering from work-stress, the additional null results give some confidence to the overall results in Figure 1. Results from those who answer 'yes, to some extent' also add confirmation to these initial findings. However, results for 'yes, definitely' are less clear. While credible interval do overlap, participant predictions appear slightly higher. However, again, the direction of any effect is difficult to disentangle.

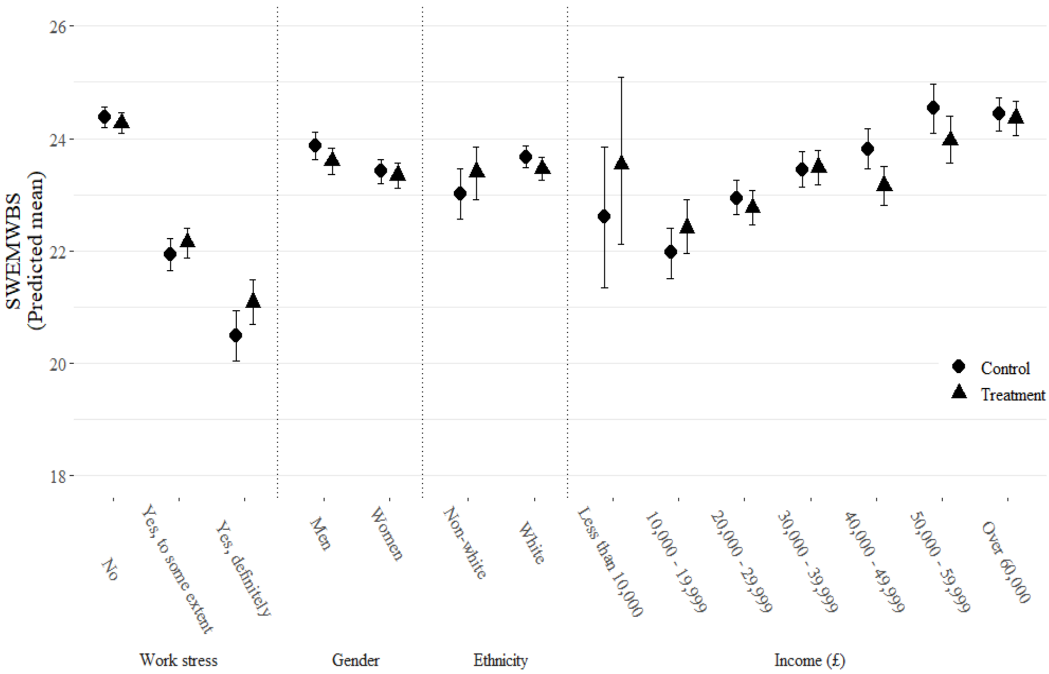
Interpreting the work-stress predictor for resilience (Figure 3) and volunteering programmes (Figure 4) also reveals a similar non-association. For resilience training there are no clear differences within the stress-level groups. For volunteering, the interaction term for work stress does not model the selection bias because the bias for the positive effect would theoretically be in the opposite direction, with happier workers more likely to participate. However, the previous stress interaction term does offer group predictions for estimating whether there is distinct benefit for those who suffer from stress. These results no longer demonstrate a positive effect, suggesting that any overall effect is small, and also that when sample size is reduced through grouping, the overall positive effect of participation I found in the primary analysis stage can no longer determined.

4.3 Job quality predictors of worker wellbeing

In light of the null and weak effects that are presented in the analysis so far, I follow existing research in suggesting that an alternative approach to workplace wellbeing is needed which considers the working environment. To provide a reference point for these debates and for possible organisational activities, this subsection presents the modelled relationships between wellbeing and dimensions of job quality. Results are in Figure 5, which includes unmatched and unweighted regression coefficients, as well as estimates for matched samples. All β coefficients are smaller for the matched models. The gap between these estimates can be explained by the increased probability of some workers reporting higher or lower values (i.e. another form of treatment of selection bias). Matching offers a more comprehensive model of confounders that predict both job quality and wellbeing, reducing the coefficients.

Figure 5 shows that all items do have a directional association with wellbeing. There are negative coefficients for workers who are bullied, face unrealistic time pressures, suffer from a work-related musculoskeletal problem, believe their organisation discriminates, and

Figure 2: Predicted SWEMWBS from interaction effects of null-effect interventions



Note: includes all interventions that estimated null effects in the previous stage of analysis: mindfulness, relaxation, coaching, time management, financial wellbeing, wellbeing app, online coaching, sleep apps, and sleep events. Whiskers show 95% highest density credible intervals.

report having strained work relationships. The largest of these estimates are for unrealistic time pressures and bullying.

All other items have positive associations. Wellbeing is a standard deviation higher for those who understand their duties and have task discretion. Flexible work arrangements like flexible hours and the ability to work from home have positive coefficients, but these are the smallest effect sizes. Having the right training, choosing break time, being consulted on change, fair pay, fair promotions, and good collaboration are all positively correlated with wellbeing.

5 Discussion

5.1 Interpreting results with selection bias

These results remain aggregative from a cross-sectional convenience sample and should only be taken as indicative evidence. Despite the optimistic epistemic claims of PSA, causal inference remains constrained by observational cross-sectional data and treatment selection bias. The strength of the PSA approach is in allowing for a more comparative sample between participants and non-participants, limiting bias at the individual level across a

Figure 3: Predicted SWEMWBS from interaction effects of resilience interventions

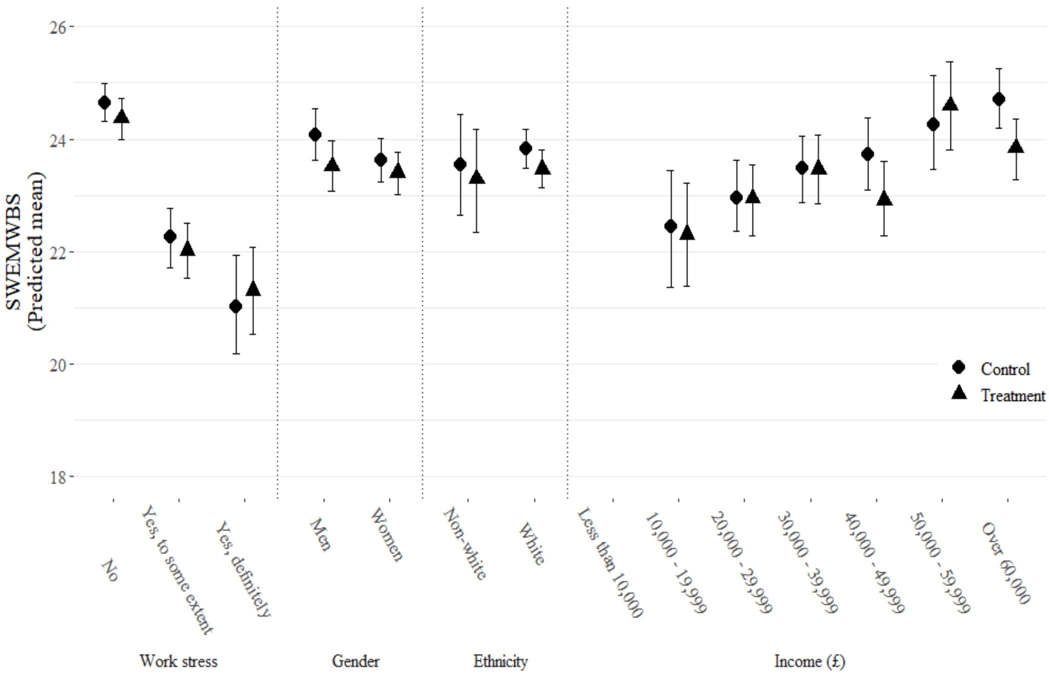
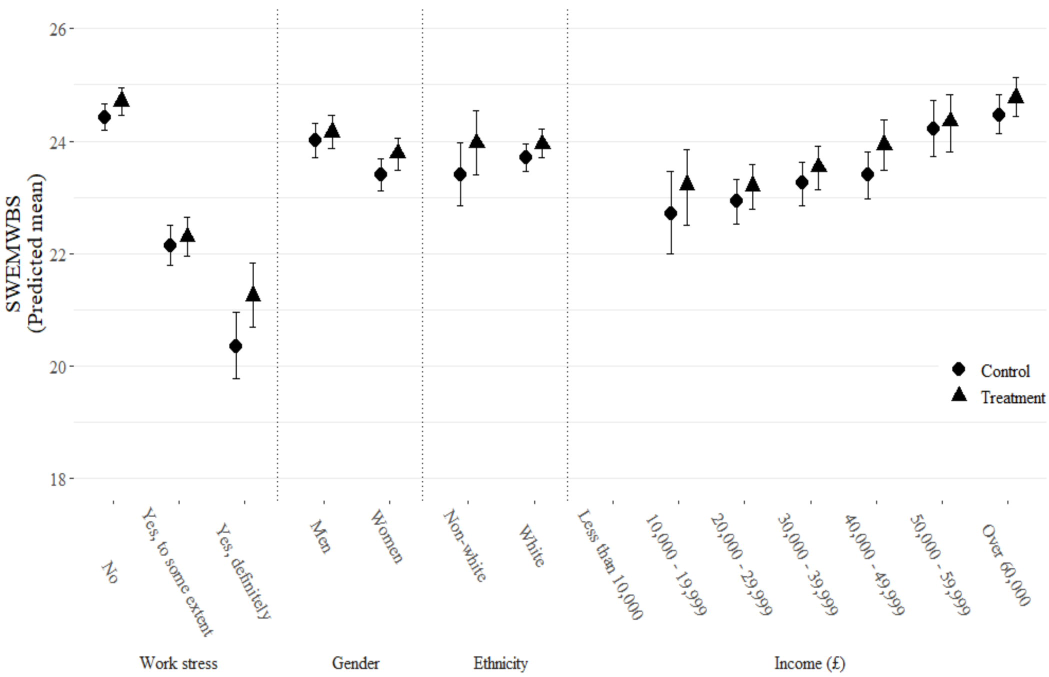
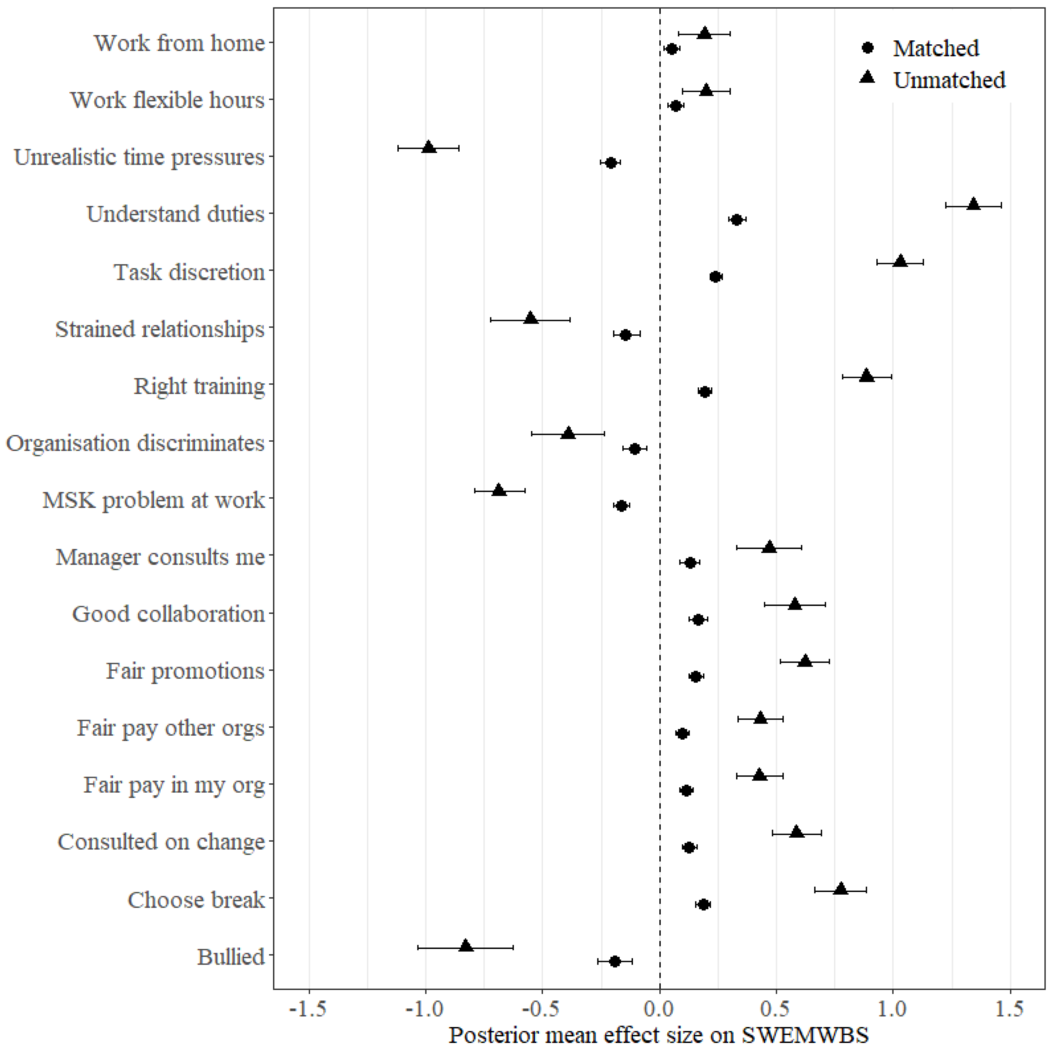


Figure 4: Predicted SWEMWBS from interaction effects of volunteering interventions



Notes for Figures 3 and 4: Whiskers show 95% highest density credible intervals. 'Less than 10,000' category excluded due to small group sample size.

Figure 5: 'Treatment effect' of workplace wellbeing dimensions on workers' SWEMWBS



Note: SWEMWBS is standardised ($\bar{y} = 0$, std.dev. = 1). β coefficient estimates are posterior mean from MCMC chains with length 5,000. Whiskers show 95% highest density credible intervals. X-axis scale differs from Figure 1.

range of observables. However, selection bias remains because the outcome, wellbeing, also predicts participation. Workers with poor wellbeing are more likely to participate (Figure A.1), but wellbeing cannot be controlled for beyond specified conditions and previous work-related stress.

For the directional relationships in resilience and volunteering programmes, it cannot be verified that the reverse causal direction is not more realistic. For resilience programmes, participation may appeal or be prescribed to specific workers in need of additional support or for specific job roles that require more support. For volunteering opportunities, it is expected that those with higher wellbeing would already be more inclined to participate as they are assumed to be more socially active. A theoretical unpacking of this remaining selection bias demands multiple interpretations of the three directional effects.

For interventions that are estimated to have no effect, there are two options: (1) there is no selection bias and there is simply no effect; or (2) there is selection bias, with those more in need of support more likely to participate, and there is a positive effect that improves wellbeing up to the level of non-participants. Option (2) implies that there is no effect from participation for those with existing good health, in that a positive effect for those with poorer prior wellbeing allows them to catch up with their happier co-workers. Considering the effect only among workers who have experienced work-related stress, it seems that there is no discernible effect from participation, giving weight to option (1). Yet regardless of which of the two possible explanations is most accurate, there would be no benefit to a universal adoption of these practices, or in adoption as a promotional strategy for improving on already high levels of wellbeing. However, these strategies may develop psychological resource in anticipation of future stress, in which case any benefit would only be detected long term, or may never be required. In the survey, a positive answer to participation indicates participation in programmes for the previous 12 months, so the answer to this suggestion remains limited.

For the negative estimates for resilience and stress management training, there are three possible explanations: (1) there is a negative effect from participating; (2) there is no effect and the negative coefficient reveals selection bias; or (3) a positive effect, but that it is small and not ameliorative enough to improve wellbeing up to the level of the control group. Options (1) and (2) would suggest that these programmes are not beneficial for workers at all and should not be implemented by employers. Explanation (3) would suggest some benefit for workers, but that as interventions they are not effective enough to compensate for poor mental health, implying that alternative strategies may be more appropriate or also required. Considering the interaction predictions in Figure 3, explanation (2) appears to be most likely. This step attempted to model whether someone was more likely to participate based on whether they experienced stress at work, also suggesting that there is no effect from participation. Whether (2) is the correct explanation or not, all three possibilities challenge the utility of resilience and stress management training.

For the positive effect of volunteering, there are also three options: (1) a positive effect; (2) no effect, with the positive estimate revealing selection bias; or (3) a positive effect that suffers from selection bias, making the true treatment effect smaller than estimated. Options (1) and (3) indicate benefits for workers, and option (2), while not benefiting the worker directly, has the positive externality of 'doing good'. All three options therefore indicate volunteering opportunities would be beneficial for workers. Volunteering also has the highest participation rates of all the programmes (Table 1), suggesting a willingness from workers to engage.

5.2 Summary of results

I have shown that those who participate in individual-level mental wellbeing interventions have the same levels of wellbeing as though who do not. This primary result is counter to much of the existing narrative in empirical literature and UK policy space. The large multi-organisational sample I used reveals that any impact on individual wellbeing is undetectable, and that this is consistent across multiple organisational contexts. While these findings do not entirely discount positive effects for some individual workers, any such effect may be averaged out by a negative effect elsewhere.

Indicative results do point towards a negative impact from participating in resilience and stress management programmes. On initial inspection, this would support Lovejoy et al. (2021) concern that these programmes imply stress is self-imposed, risking negative impact, and also Frayne's (2019) argument that resilience training is an attempt to adapt workers to workplace stress, rather than alleviating or preventing it, and that this can cause stress itself as a 'toxic side effect'. However, once selection bias is modelled by considering whether workers had been negatively affected by work-related stress, this negative effect is flattened, suggesting consistency with the other null results. The positive effect of volunteering opportunities is the exception to the overall findings. However, this possible benefit is estimated to be small, especially considering that the positive effect is no longer clear when grouped treatment effects are predicted.

5.3 Strengths and further limitations

There are further limitations to this analysis beyond the obvious methodological problems of cross-sectional rather than longitudinal data, as well as the discussed treatment selection bias. I have only evaluated the interventions in isolation, meaning that, although the organisation context and some organisational-level variables are controlled for, participation is not considered within the wider structures that are important for evaluation (Daniels et al., 2022). Further, little information on the interventions is provided by the BHW; there will be differences in what service was provided. This is particularly relevant for the estimates of volunteering opportunities, as options may be explicit community initiatives, or may be pro-bono professional work. Finally, information on why people have participated would allow for a more complete modelling of selection bias. Greater survey coverage is required to address most of these empirical limitations.

Although most weaknesses of this analysis arise from the data used, the scale and coverage of the same data also offers the main strength of this article. The BHW survey provides a sample of British workers larger than any previous study evaluating interventions. It also has a large number of organisations, allowing the incorporation of group and industry-level variance to statistical and theoretical models. The range of interventions I have evaluated is more comprehensive than any of the single programme RCTs that constitute the existing evidence base. While experimental studies will provide more compelling causal evidence, their scale is limited practically. Further, in examining multiple interventions across multiple organisations, my analysis will include varying degrees of fidelity in interventions; a pressing need in the existing literature (Fikretoglu et al., 2022) and relevant to the real-life implementation of wellbeing interventions. RCTs and other implementation studies ensure high fidelity in the deployment of an intervention – i.e. ensuring that it is implemented as intended in a standardised way – to accurately determine its effects. It seems reasonable to suggest that examining a range of organisations with observational data like the BHW likely includes initiatives that are not so well implemented. This offers another possible explanation for the null results.

Further, I have used multiple subjective wellbeing outcomes to evaluate the mental wellbeing interventions as health initiatives, not just business practices. Doing so ensures that that this article provides supporting observational evidence for both systematic reviews of RCTs and, by using provocations from critical sociological accounts of workplace wellbeing, begins to theorise the causal mechanisms of these strategies, contributing to sociological, management and public health literature on workplace wellbeing. While limited, it is hoped that these findings can still make a worthwhile contribution to the broader empirical and normative debates.

To build on both these strengths and weaknesses, further research into labour relations and workers' experience of wellbeing interventions is required. There is much discursive work in critical accounts and some in-depth qualitative research, but there remains a need for qualitative evaluation and the inclusion of worker voice in discussions. Qualitative research would provide a greater understanding of the diverse contexts in which interventions are implemented. Finally, further longitudinal and experimental studies should be undertaken to better establish potential positive and negative effects. To enhance external validity, these studies must be large scale and in multiple contexts, as well as pay particular attention to the limitations of these approaches, such as continued treatment selection bias, attrition, and the causes of both.

6 Conclusion

The results I share in this article pose a challenge to the popularity and legitimacy of individual-level mental wellbeing interventions like mindfulness, resilience and stress management, relaxation classes, and wellbeing apps. I find little evidence in support of any benefits from these interventions. Additionally, there is even some indication of harm, which would confirm fears from critics.

The overall findings must also be placed in sociological debates on workplace wellbeing practices, which see these practices as potentially depoliticising, exploitative, or simply strategic CSR. These results cannot verify or falsify such ideological critiques, but they do supplement arguments challenging the empirical legitimacy of promotional, individual-level strategies that appear to operate in isolation from working conditions.

The results also echo existing empirical studies that seek to understand the effect on workers, where low-level interventions such as these are not found to be effective for wellbeing unless linked to preventative structural change (Daniels et al., 2021b; LaMontagne et al., 2007), or where no effect, accompanied by treatment selection bias, is found in wellness programmes (Jones et al., 2019; Song & Baicker, 2019). Finally, they also contribute to the study of 'what works', as well as expanding this question to consider 'for whom and in what circumstances' (Nielsen & Miraglia, 2017). In reviewing the literature, I have also highlighted ongoing limitations in the existing evidence base and so asserting the best approach for improving workers' wellbeing is an ongoing endeavour.

In the absence of further clarity on how these interventions would benefit workers and on the working contexts they are implemented in, I argue that they should not be recommended as a one-size-fits-all solution for improving workers' wellbeing. In formal recommendations from NICE, individualised approaches are only recommended alongside organisational change. It seems premature to advise strategies like mindfulness as universal strategies for improving and maintain worker mental health. With optimistic claims of improved wellbeing and productivity, managers may follow the evidence in the belief they are operating in the best interests of workers' wellbeing. However, regardless of intention, the poor existing evidence base and these new results undermines this 'win-win' rhetoric at the centre of individualised wellbeing strategies. Together, they suggest that merely offering short-term 'off-the-shelf' programmes or classes will not prove a satisfactory solution for the systemic problems in working conditions and work-induced stress, both of which are said to be intensifying (e.g. Kalleberg, 2011; Kelly & Moen, 2020). In agreement with other researchers, I recommend wellbeing interventions which seek organisational change and that engage with working conditions.

The positive effect of volunteering opportunities offers exception to some of the commentary included in this discussion section, and does offers some positive news. Remaining selection bias has been discussed, but results would echo much experimental evidence. Volunteering is increasingly advocated in individual mental wellbeing plans and promoted by policy think-tanks on similar terms (e.g. Stuart et al., 2020). The finding that volunteering opportunities are good for workers' wellbeing suggests that more tied to civic society, providing more meaningful work. It also suggests an alternative win-win model of wellbeing that is more collective and relational (Atkinson, 2021). This result then provides support for the adoption of volunteering opportunities for workers, regardless of biased estimates. However, these strategies will still not address working conditions and the causes of work-related stress, and so remain only one component of a necessary organisational strategy.

*

References

Aas, R. W., Tuntland, H., Holte, K. A., Røe, C., Lund, T., Moller, A., & Marklund, S. (2011). Workplace interventions for neck pain in workers. *Cochrane Database of Systematic Reviews*, *4*.

Atkinson, S. (2021). The toxic effects of subjective wellbeing and potential tonics. *Social Science & Medicine*, *288*(113098).

Baicker, K., Cutler, D., & Song, Z. (2010). Workplace Wellness Programs Can Generate Savings. *Health Affairs*, *29*(2): 304–311.

Bartlett, L., Martin, A., Neil, A. L., Memish, K., Otahal, P., Kilpatrick, M., & Sanderson, K. (2019). A systematic review and meta-analysis of workplace mindfulness training randomized controlled trials. *Journal of Occupational Health Psychology*, *24*(1): 108–126.

Black, C. (2008). Working for a healthier tomorrow. Department for Work and Pensions; Department of Health, London.
URL https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/209782/hwwb-working-for-a-healthier-tomorrow.pdf

Bostock, S., Crosswell, A. D., Prather, A. A., & Steptoe, A. (2019). Mindfulness on-the-go: Effects of a mindfulness meditation app on work stress and well-being. *Journal of Occupational Health Psychology*, *24*(1): 127–138.

Burton, N. W., Pakenham, K. I., & Brown, W. J. (2010). Feasibility and effectiveness of psychosocial resilience training: A pilot study of the *READY* program. *Psychology, Health & Medicine*, *15*(3): 266–277.

Cannas, M. (2019). CMatching: Matching Algorithms for Causal Inference with Clustered Data.
URL https://CRAN.R-project.org/package=CMatching

CIPD (2022). Health and Wellbeing at Work 2022. Chartered Institute of Personnel and Development, London.

Creswell, J. D. (2017). Mindfulness Interventions. *Annual Review of Psychology*, *68*(1): 491–516.

Daniels, K., Fida, R., Stepanek, M., & Gendronneau, C. (2021a). Do Multicomponent Workplace Health and Wellbeing Programs Predict Changes in Health and Wellbeing? *International Journal of Environmental Research and Public Health*, *18*(17).

Daniels, K., Tregaskis, O., Nayani, R., & Watson, D. (2022). *Achieving Sustainable Workplace Wellbeing*. Cham: Springer.

Daniels, K., Watson, D., Nayani, R., Tregaskis, O., Hogg, M., Etuknwa, A., & Semkina, A. (2021b). Implementing practices focused on workplace health and psychological wellbeing: A systematic review. *Social Science & Medicine*, *277*: 113888.

Davies, W. (2015). *The Happiness Industry: How the Government and Big Business Sold Us Well-Being*. London: Verso Books.

DWP, & DHSC (2017). Improving Lives: The Future of Work, Health and Disability. Department for Work and Pensions; Department of Health and Social Care, London. URL `https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/663399/improving-lives-the-future-of-work-health-and-disability.PDF`

Eisendrath, S. J., Gillung, E., Delucchi, K. L., Segal, Z. V., Nelson, J. C., McInnes, L. A., Mathalon, D. H., & Feldman, M. D. (2016). A Randomized Controlled Trial of Mindfulness-Based Cognitive Therapy for Treatment-Resistant Depression. *Psychotherapy and psychosomatics*, *85*(2): 99–110.

Fida, R., Game, A., Stepanek, M., & Gendronneau, C. (2021). Longitudinal Effects of Engagement with Workplace Health Programmes on Employee Outcomes: A Relational Perspective. *British Journal of Management*, *Early View*.

Fikretoglu, D., Easterbrook, B., & Nazarov, A. (2022). Fidelity in workplace mental health intervention research: A narrative review. *Work & Stress*, *36*(1): 6–29.

Foster, D. (2018). The Health and Well-Being at Work Agenda: Good News for (Disabled) Workers or Just a Capital Idea? *Work, Employment and Society*, *32*(1): 186–197.

Fox, K. E., Johnson, S. T., Berkman, L. F., Sianoja, M., Soh, Y., Kubzansky, L. D., & Kelly, E. L. (2022). Organisational- and group-level workplace interventions and their effect on multiple domains of worker well-being: A systematic review. *Work & Stress*, *36*(1): 30–59.

Frayne, D. (Ed.) (2019). *The Work Cure: Critical Essays on Work and Wellness*. Monmouth: PCCS Books.

Galante, J., Stochl, J., Dufour, G., Vainre, M., Wagner, A. P., & Jones, B. P. (2020). Effectiveness of providing university students with a mindfulness-based intervention to increase resilience to stress: 1-year follow-up of a pragmatic randomised controlled trial. *Journal of Epidemiology & Community Health*, *Online First*.

Goetzel, R. Z. (2020). Commentary on the Study: "What Do Workplace Wellness Programs Do? Evidence From the Illinois Workplace Wellness Study". *American Journal of Health Promotion*, *34*(4): 440–444.

Goldberg, S. B., Riordan, K. M., Sun, S., & Davidson, R. J. (2021). The Empirical Status of Mindfulness-Based Interventions: A Systematic Review of 44 Meta-Analyses of Randomized Controlled Trials. *Perspectives on Psychological Science*, *OnlineFirst*.

Hampson, E., Soneji, U., Jacob, A., Mecu, B., & McGahan, G. (2017). Mental health and employers: The case for investment. Monitor Deloitte, London.

Hargrave, G. E., & Hiatt, D. (2005). The EAP Treatment of Depressed Employees: Implications for Return on Investment. *Employee Assistance Quarterly*, *19*(4): 39–49.

Hoge, E. A., Bui, E., Marques, L., Metcalf, C. A., Morris, L. K., Robinaugh, D. J., Worthington, J. J., Pollack, M. H., & Simon, N. M. (2013). Randomized Controlled Trial of Mindfulness Meditation for Generalized Anxiety Disorder: Effects on Anxiety and Stress Reactivity. *The Journal of clinical psychiatry*, *74*(8): 786–792.

Holmqvist, M., & Maravelias, C. (2011). *Managing Healthy Organizations: Worksite Health Promotion and the New Self-Management Paradigm*. Abingdon: Routledge.

Hood, B., Jelbert, S., & Santos, L. R. (2021). Benefits of psychoeducational happiness course on university student mental well-being both before and during a COVID-19 lockdown. *Health Psychology Open*, *8*(1).

Hull, G., & Pasquale, F. (2018). Toward a critical theory of corporate wellness. *BioSocieties*, *13*(1): 190–212.

Islam, G., Holm, M., & Karjalainen, M. (2022). Sign of the times: Workplace mindfulness as an empty signifier. *Organization*, *29*(1): 3–29.

Jones, D., Molitor, D., & Reif, J. (2019). What do Workplace Wellness Programs do? Evidence from the Illinois Workplace Wellness Study. *The Quarterly Journal of Economics*, *134*(4): 1747–1791.

Joyce, S., Shand, F., Tighe, J., Laurent, S. J., Bryant, R. A., & Harvey, S. B. (2018). Road to resilience: A systematic review and meta-analysis of resilience training programmes and interventions. *BMJ Open*, *8*(6).

Kalleberg, A. L. (2011). *Good Jobs, Bad Jobs: The Rise of Polarized and Precarious Employment Systems in the United States, 1970s to 2000s*. New York: Russell Sage.

Kaplan, D., & Chen, J. (2012). A Two-Step Bayesian Approach for Propensity Score Analysis: Simulations and Case Study. *Psychometrika*, *77*: 581–609.

Kelly, E. L., & Moen, P. (2020). *Overload: How Good Jobs Went Bad and What We Can Do About It*. Princeton, NJ: Princeton University Press.

King, G., & Nielsen, R. (2019). Why Propensity Scores Should Not Be Used for Matching. *Political Analysis*, *27*(4): 435–454.

Knapp, M., McDaid, D., & Parsonage, M. (2011). Mental health promotion and mental illness prevention: The economic case. Department of Health, London.

Kowalski, T. H., & Loretto, W. (2017). Well-being and HRM in the changing workplace. *The International Journal of Human Resource Management*, *28*(16): 2229–2255.

LaMontagne, A. D., Keegel, T., Louie, A. M., Ostry, A., & Landsbergis, P. A. (2007). A Systematic Review of the Job-stress Intervention Evaluation Literature, 1990-2005. *International Journal of Occupational and Environmental Health*, *13*(3): 268–280.

Lovejoy, M., Kelly, E. L., Kubzansky, L. D., & Berkman, L. F. (2021). Work Redesign for the 21st Century: Promising Strategies for Enhancing Worker Well-Being. *American Journal of Public Health*, *111*(10): 1787–1795.

Matrix (2013). Economic analysis of workplace mental health promotion and mental disorder prevention programmes and of their potential contribution to EU health, social and economic policy objectives. Matrix Knowledge, London.
URL `https://ec.europa.eu/health/sites/health/files/mental_health/docs/matrix_economic_analysis_mh_promotion_en.pdf`

Milligan-Saville, J. S., Tan, L., Gayed, A., Barnes, C., Madan, I., Dobson, M., Bryant, R. A., Christensen, H., Mykletun, A., & Harvey, S. B. (2017). Workplace mental health training for managers and its effect on sick leave in employees: A cluster randomised controlled trial. *The Lancet: Psychiatry*, *4*(11): 850–858.

Mills, P. R., Kessler, R. C., Cooper, J., & Sullivan, S. (2007). Impact of a health promotion program on employee health risks and work productivity. *American Journal of Health Promotion*, *22*(1): 45–53.

Murphy, C., & MacMahon, J. (2022). Employee resilience: A labour process perspective. *Work in the Global Economy*, *Early View*.

Newman, R. (2021). Evidence-based approaches to workplace mental health.
URL `https://whatworkswellbeing.org/blog/evidence-based-approaches-to-workplace-mental-health/`

NICE (2022). Mental wellbeing at work. National Institute for Health and Care Excellence.

Nielsen, K., & Miraglia, M. (2017). What works for whom in which circumstances? On the need to move beyond the 'what works?' question in organizational intervention research. *Human Relations*, *70*(1): 40–62.

Noordik, E., van der Klink, J. J., Klingen, E. F., Nieuwenhuijsen, K., & van Dijk, F. J. (2010). Exposure-in-vivo containing interventions to improve work functioning of workers with anxiety disorder: A systematic review. *BMC Public Health*, *10*(1): 598.

Pipe, T. B., Buchda, V. L., Launder, S., Hudak, B., Hulvey, L., Karns, K. E., & Pendergast, D. (2012). Building Personal and Professional Resources of Resilience and Agility in the Healthcare Workplace. *Stress and Health*, *28*(1): 11–22.

Pomaki, G., Franche, R.-L., Murray, E., Khushrushahi, N., & Lampinen, T. M. (2012). Workplace-Based Work Disability Prevention Interventions for Workers with Common Mental Health Conditions: A Review of the Literature. *Journal of Occupational Rehabilitation*, *22*(2): 182–195.

PWC (2014). Creating a mentally healthy workplace: Return on investment analysis. URL https://www.pwc.com.au/publications/pdf/beyondblue-workplace-roi-may14.pdf

Robertson, I. T., Cooper, C. L., Sarkar, M., & Curran, T. (2015). Resilience training in the workplace from 2003 to 2014: A systematic review. *Journal of Occupational and Organizational Psychology*, *88*(3): 533–562.

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, *70*(1): 41–55.

Rubin, D. B. (1997). Estimating Causal Effects from Large Data Sets Using Propensity Scores. *Annals of Internal Medicine*, *127*(8): 757–763.

Schaufeli, W., & Bakker, A. (2004). Utrecht Work Engagement Scale: Preliminary Manual. Occupational Health Psychology Unit, Utrecht University, Utrecht. URL https://www.wilmarschaufeli.nl/publications/Schaufeli/Test%20Manuals/Test_manual_UWES_English.pdf

Sherlock-Storey, M., Moss, M., & Timson, S. (2013). Brief coaching for resilience during organisational change - an exploratory study. *Coaching Psychologist*, *9*(1): 19–26.

Song, Z., & Baicker, K. (2019). Effect of a Workplace Wellness Program on Employee Health and Economic Outcomes: A Randomized Clinical Trial. *Journal of American Medical Association*, *321*(15): 1491–1501.

Sood, A., Prasad, K., Shroeder, D., & Varkey, P. (2011). Stress Management and Resilience Training Among Department of Medicine Faculty: A Pilot Randomized Clinical Trial. *Journal of General Internal Medicine*, *26*: 858–861.

Sorensen, G., Dennerlein, J. T., Peters, S. E., Sabbath, E. L., Kelly, E. L., & Wagner, G. R. (2021). The future of research on work, safety, health and wellbeing: A guiding conceptual framework. *Social Science & Medicine*, *269*(113593).

Southwood, I. (2019). The black dog. In D. Frayne (Ed.) *The Work Cure: Critical Essays on Work and Wellness*, (pp. 29–44). Monmouth: PCCS Books.

Stevenson, D., & Farmer, P. (2017). Thriving at Work: The Stevenson/Farmer review of mental health and employers. Department for Work and Pensions; Department of Health and Social Care, London. URL https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/658145/thriving-at-work-stevenson-farmer-review.pdf

Stuart, J., Kamerāde, D., Connolly, S., Ellis Paine, A., Nichols, G., & Grotz, J. (2020). The Impacts of Volunteering on the Subjective Wellbeing of Volunteers: A Rapid Evidence Assessment. What Works Centre for Wellbeing.

Tetrick, L. E., & Winslow, C. J. (2015). Workplace Stress Management Interventions and Health Promotion. *Annual Review of Organizational Psychology and Organizational Behaviour*, *2*: 583–603.

TUC (2013). Work and well-being: A trade union resource. Trade Union Congress, London.

TUC (2018). Mental health and the workplace. Trade Union Congress, London.

Tveito, T. H., Hysing, M., & Eriksen, H. R. (2004). Low back pain interventions at the workplace: A systematic literature review. *Occupational Medicine*, *54*(1): 3–13.

Veld, M., & Alfes, K. (2017). HRM, climate and employee well-being: Comparing an optimistic and critical perspective. *The International Journal of Human Resource Management*, *28*(16): 2299–2318.

Vonderlin, R., Biermann, M., Bohus, M., & Lyssenko, L. (2020). Mindfulness-Based Programs in the Workplace: A Meta-Analysis of Randomized Controlled Trials. *Mindfulness*, *11*(7): 1579–1598.

Wagner, S., Koehn, C., White, M., Harder, H. G., Schultz, I., Williams-Whitt, K., Wärje, O., Dionne, C., Koehoorn, M., Pasca, R., Hsu, V., McGuire, L., Schulz, W., Kube, D., & Wright, M. (2016). Mental Health Interventions in the Workplace and Work Outcomes: A Best-Evidence Synthesis of Systematic Reviews. *The International Journal of Occupational and Environmental Medicine*, *7*(1): 1–14.

Wallace, J. (2019). *Productive Sickness: Wellbeing Discourse, Employee Subjectivity and the Organisation of Ill-Health*. Ph.D. thesis, University of Cardiff.

Wang, P. S., Simon, G. E., Avorn, J., Azocar, F., Ludman, E. J., McCulloch, J., Petukhova, M. Z., & Kessler, R. C. (2007). Telephone Screening, Outreach, and Care Management for Depressed Workers and Impact on Clinical and Work Productivity Outcomes: A Randomized Controlled Trial. *JAMA*, *298*(12): 1401–1411.

Western, B. (1999). Bayesian Analysis for Sociologists: An Introduction. *Sociological Methods & Research*, *28*(1): 7–34.

Zhang, Z., Parker, R. M., Charlton, C. M., Leckie, G., & Browne, W. J. (2016). R2MLwiN: A Package to Run MLwiN from within R. *Journal of Statistical Software*, *72*(10): 1–43.

Zoller, H. M. (2004). Manufacturing health: Employee perspectives on problematic outcomes in a workplace health promotion initiative. *Western Journal of Communication*, *68*(3): 278–301.

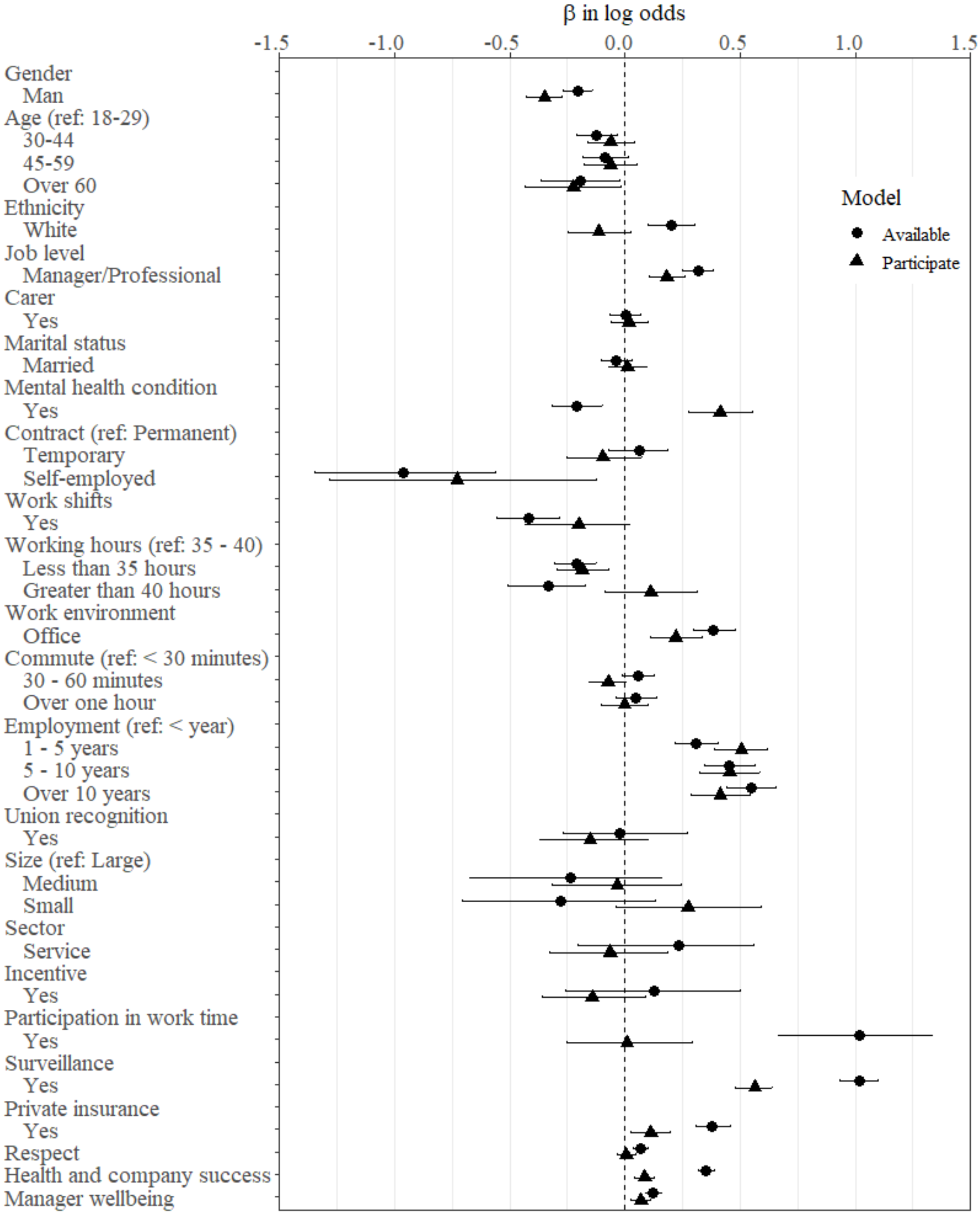
A Appendix

Table A.1: List of interventions and post-matching sample sizes

Intervention	Employees N	Organisations N
Volunteering or charity work	4,532	103
Mindfulness classes or programmes	2,556	89
Resilience, energy or stress management classes or programmes	1,756	80
Wellbeing app for physical health, mental health and lifestyle issues	1,562	78
Massage or relaxation classes or programmes	1,510	79
Workload or time management training	988	85
Financial wellbeing courses or programmes	638	55
Coaching (one-on-one sessions on mental health and wellbeing)	358	53
Apps/programmes promoting healthy sleep	346	31
Events promoting healthy sleep	322	32
Online coaching	156	28
All mental health promotion interventions	9.544	123

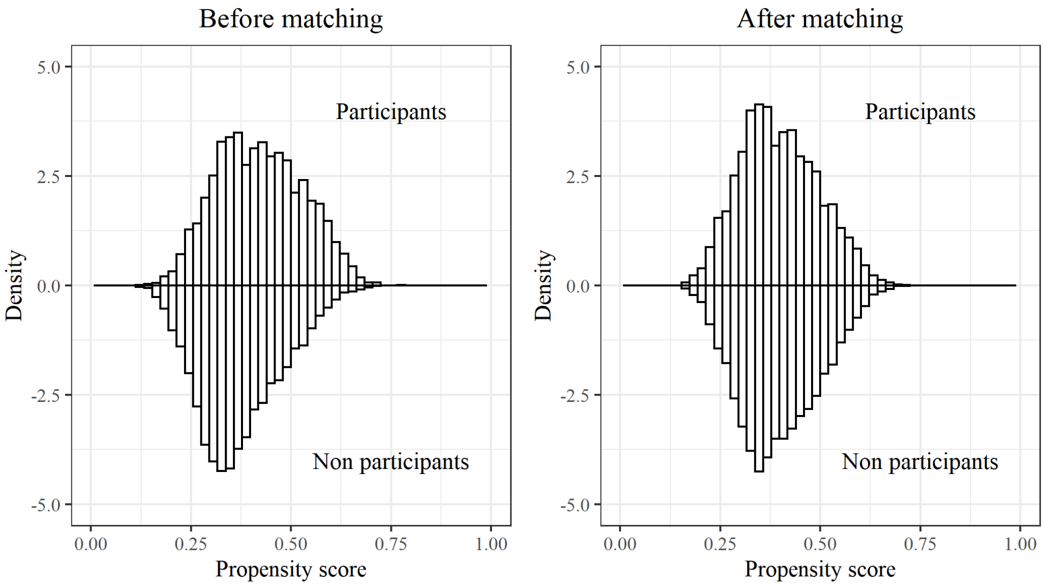
Note: Sample sizes for each type of intervention following the matching process outlined in the analytical strategy. Full sample: employees N = 27,919; organisations N = 143. Sorted by Employees N.

Figure A.1: Coefficient plot of mental health promotion models



Note: β coefficients reported with 95% credible intervals. 'Respect', 'health and company success', and 'manager wellbeing' variables are all standardised with mean 0 and std. dev. 1.

Figure A.2: Propensity score density plots for participation in ‘any’ interventions



Note: Figure shows density plots of BHW individuals’ probability of participation, propensity scores, for participants and non-participants before and after matching. Before matching reveals selection bias among participants whereas after matching provides a balanced sample of propensity scores. Figure shows for all interventions combined but the process is undertaken for each interventions separately.

Table A.2: BHW questions on work environment

Question	Yes	No
'*Staff are always consulted on organisational change*'	Ag, SA	Neu, SD, D
'*I have the training and tools I need to do my job*'	Ag, SA	Neu, D, SD
'*Are you able to work flexible hours?*'	Yes	No
'*Are you able to work from home (at least some of the time)?*'	Yes	No
'*I have unrealistic time pressures*'	O, Al	So, Se, Nev
'*I have a choice in deciding what I do at work*'	O, Al	So, Se, Nev
'*I am clear what my duties and responsibilities are*'	O, Al	So, Se, Nev
'*I can decide to take a break during the working day*'	O, Al	So, Se, Nev
'*Promotions and salary decisions in my organisation are fair and transparent*'	A, SA	Neu, D, SD
'*I am fairly paid in comparison to similar people in similar roles within the organisation*'	Ag, SA	Neu, D, SD
'*I am fairly paid in comparison to similar people in similar roles in other organisations*'	Ag, SA	Neu, D, SD
'*I am subject to bullying at work*'	So, O, Al	Nev, Se
'*Do you think any of the pain or discomfort is a result of your workplace conditions*'	Yes	No, DK
'*Relationships at work are strained*'	A, SA	Neu, D, SD
'*There is good collaboration between staff*'	Ag, SA	Neu, D, SD
'*My line manager consults me on matters of importance to me*'	Ag, SA	Neu, D, SD
'*My organisation does not discriminate on the grounds of* [age, disability, gender, maternity, race, religion, sexuality, transgender]'	Neu, D, SD	Ag, SA

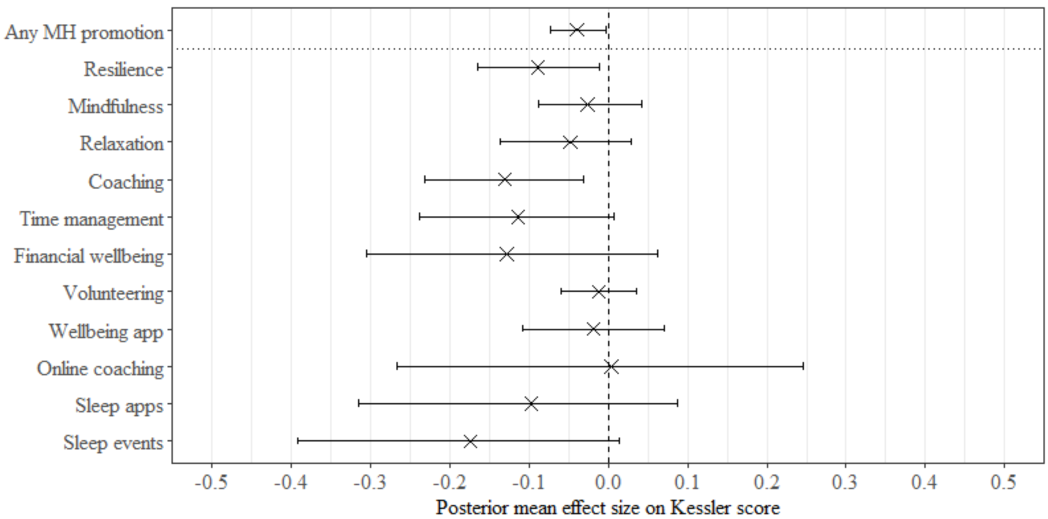
Abbreviations: SD = Strongly Disagree; D = Disagree; Neu = Neutral; Ag = Agree; SA = Strongly Agree. Nev = Never; Se = Seldom; So = Sometimes; O = Often; Al = Always. DK = Don't know.

Table A.3: Descriptive statistics of SWEMWBS outcomes and random slope covariance

Intervention	Tr mean	Co mean	Tr median	Co median	Tr sd	Co sd	σ^2_{u1}
Any MH promotion	23.41	23.35	23.21	23.21	3.90	3.89	0.003
Resilience	23.40	23.78	23.21	24.11	3.87	4.02	0.029
Mindfulness	23.48	23.58	23.21	23.21	3.93	3.93	0.002
Relaxation	23.74	23.74	24.11	24.11	4.03	3.79	0.004
Coaching	23.95	23.96	24.11	23.66	4.17	3.97	0.003
Time management	24.07	24.19	24.11	24.11	4.36	4.10	0.022
Financial wellbeing	23.56	23.96	23.21	24.11	3.97	4.03	0.008
Volunteering	23.73	23.46	24.11	23.21	3.94	3.85	0.002
Wellbeing app	23.79	23.66	23.21	24.11	4.37	4.06	0.005
Online coaching	24.40	24.46	24.11	24.11	4.63	4.52	0.007
Sleep apps	23.75	23.51	24.11	23.21	4.41	4.24	0.009
Sleep events	23.71	23.62	24.11	23.21	4.40	4.23	0.004

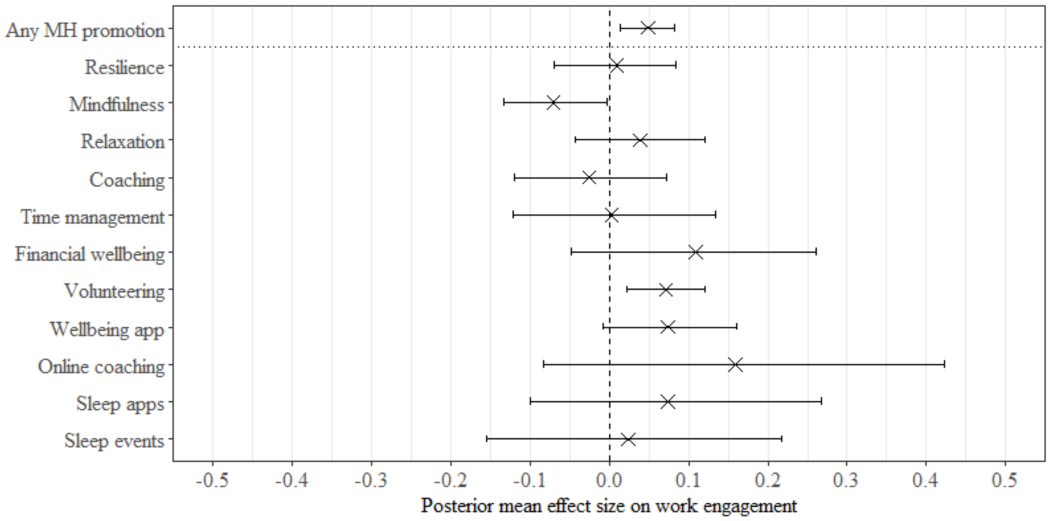
Note: Tr = treatment; Co = control; sd = standard deviation; σ^2_{u1} *= random slopes variance.*

Figure A.3: 'Treatment effect' of workplace mental wellbeing interventions on workers' Kessler scores



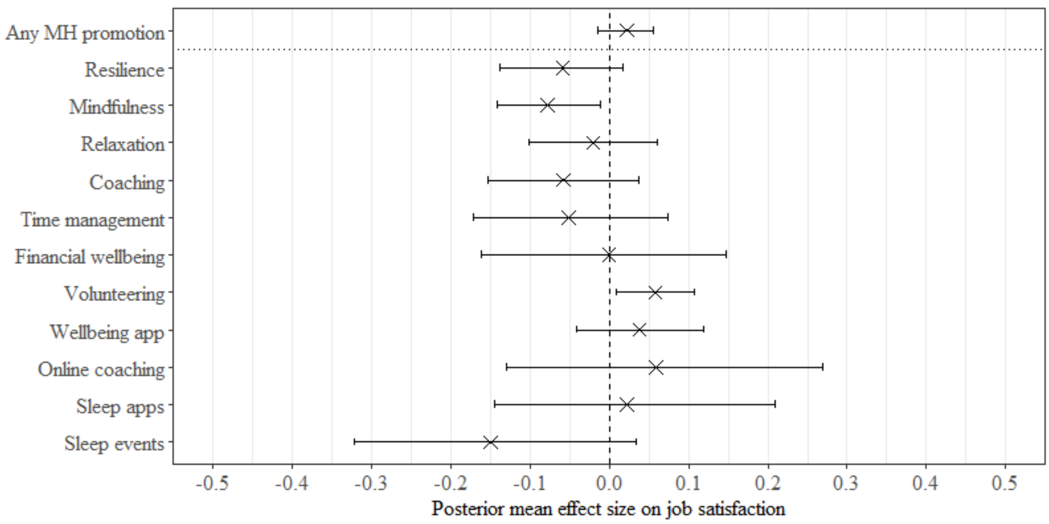
Note: Kessler is standardised ($\bar{y}$ = 0, std.dev. = 1). β Coefficient estimates are posterior mean from MCMC chains with length 5,000. Whiskers show 95% highest density credible intervals.

Figure A.4: 'Treatment effect' of workplace mental wellbeing interventions on workers' work engagement scores



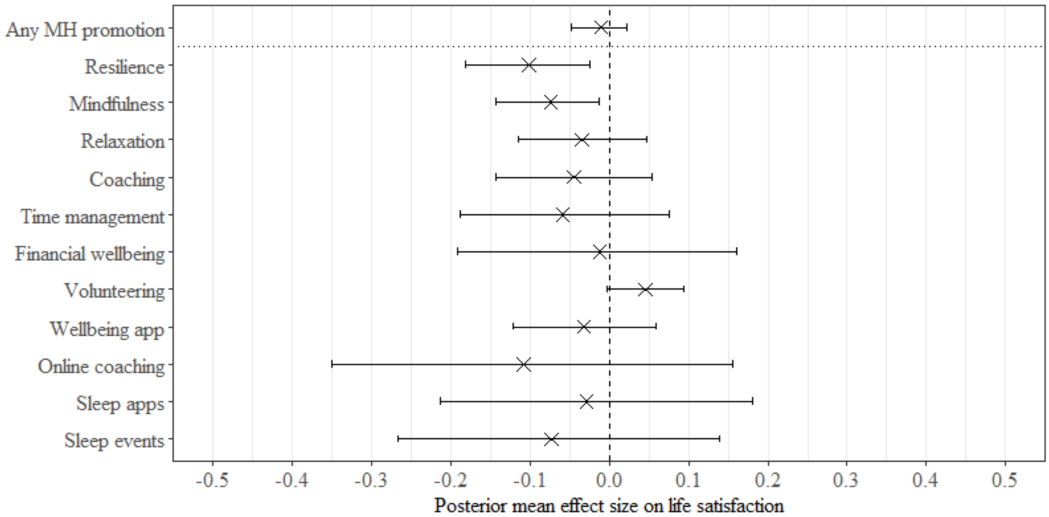
Note: Work engagement is standardised ($\bar{y} = 0$, std.dev. $= 1$). β Coefficient estimates are posterior mean from MCMC chains with length 5,000. Whiskers show 95% highest density credible intervals.

Figure A.5: 'Treatment effect' of workplace mental wellbeing interventions on workers' job satisfaction scores



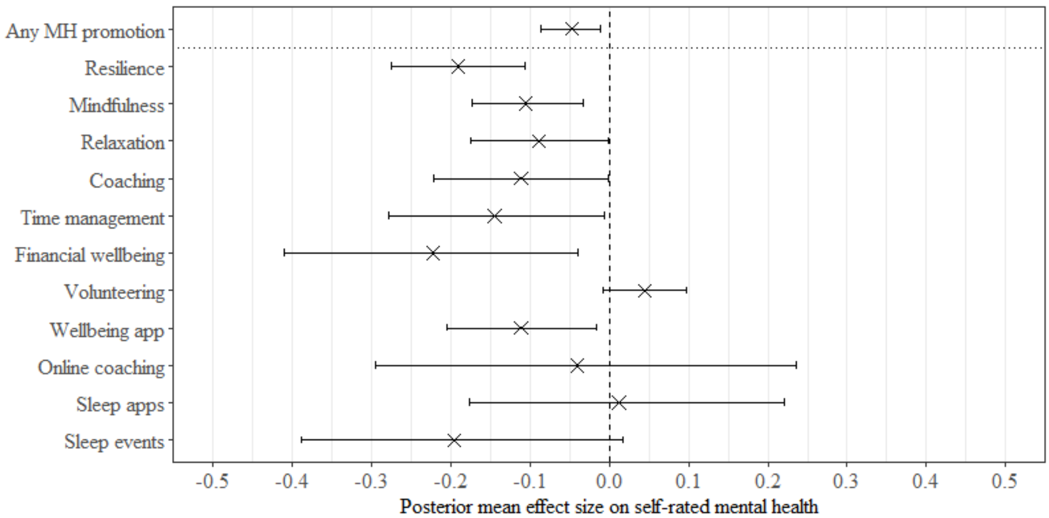
Note: Job satisfaction is standardised ($\bar{y} = 0$, std.dev. $= 1$). β Coefficient estimates are posterior mean from MCMC chains with length 5,000. Whiskers show 95% highest density credible intervals.

Figure A.6: 'Treatment effect' of workplace mental wellbeing interventions on workers' life satisfaction scores



Note: Life satisfaction is standardised ($\bar{y}$ = 0, std.dev. = 1). β Coefficient estimates are posterior mean from MCMC chains with length 5,000. Whiskers show 95% highest density credible intervals.

Figure A.7: 'Treatment effect' of workplace mental wellbeing interventions on workers' self-reported mental health



Note: Mental health is standardised ($\bar{y}$ = 0, std.dev. = 1). β Coefficient estimates are posterior mean from MCMC chains with length 5,000. Whiskers show 95% highest density credible intervals.