In this appendix, we use causal diagrams to expand on how stressors measured in different domains may confound, and perhaps also partially mediate, each other’s effects on psychological distress. Causal diagrams, also known as Directed Acyclic Graphs, are now a core tool for analyzing bias in the epidemiological literature. Readers unfamiliar with them are referred to references [1,2] for accessible introductions.

Figure A1.1 is a causal diagram representing our conceptual relationship between stressors in two different domains and psychological distress over time. For simplification, the figure only includes one (perceived) work pressure stressor variable, W; one (perceived) life stressor variable, L; and the psychological distress variable, D. The subscripts represent time, so that \( W_0 \) is the work stressor level in arbitrary time period 0, \( W_1 \) the work stressor level in arbitrary time period 1, and so on. Note that additional domains, such as stressors related to work conflict, could be added without changing the principles discussed here.

The arrows show our assumed causal relationships between variables. In Figure A1, we assume that both work pressure and life stressors in any given time period cause psychological distress (\( W_0 \rightarrow D_0 \) and \( L_0 \rightarrow D_0 \)) and that psychological distress subsequently affects the level of work pressure and life stressors in the next time period (\( D_0 \rightarrow W_1 \) and \( D_0 \rightarrow L_1 \)). For example, we can imagine that a private-life stressor such as conflict with spouse increases an individual’s distress, which then increases the likelihood of conflict with colleagues in the workplace. This shows graphically the phenomenon of reverse causation recognized in the job-stress literature [3].
We can use this diagram to identify confounding pathways for a given association of interest, assuming of course that the depicted relationships are correct. We wish to estimate the association between $W_2$ and $D_2$ in a cross-sectional survey at time period 2. There are confounding pathways linking $W_2$ to $D_2$ other than the direct path $W_2 \rightarrow D_2$, e.g. $W_2 \leftarrow D_1 \rightarrow L_2 \rightarrow D_2$, $W_2 \leftarrow W_1 \rightarrow D_1 \rightarrow L_2 \rightarrow D_2$, and so on. These are known in the literature as “unblocked backdoor” paths [1] and represent pathways confounding or biasing the estimation of the effect of work pressure stressors in the current period, $W_2$, on distress in the current period, $D_2$. (See [1,2] for a detailed explanation of how to identify confounding pathways using causal diagrams, including the definition of an open backdoor path which includes some details unable to be discussed here.)

Taking a concrete example, we can explain these confounding paths as follows: employees with psychological distress in the time period of the survey ($D_2$) have more work pressure stressors ($W_2$) but they also have more private-life stressors ($L_2$). In particular, the employees with more work pressure stressors in the survey ($W_2$) are more likely to have more private life stressors in survey ($L_2$) in part because they had higher distress in the previous time period ($D_1$). The distribution of private-life stressors ($L_2$) is therefore associated with that of work pressure stressors ($W_2$) in part because of previous psychological distress ($D_1$), confounding the observed association of work pressure stressors ($W_2$) on psychological distress levels ($D_2$) in the current period.

Confounding can be removed by statistically adjusting on a variable lying on these open backdoor paths [1]. In Figure 1, adjusting on $L_2$, e.g. by including it in a regression model, would block most of these pathways. Nonetheless, open backdoor paths involving $D_1 \rightarrow D_2$ (e.g. $W_2 \leftarrow W_1 \leftarrow D_1 \rightarrow D_2$) would remain and, if quantitatively important, would create residual
confounding. Blocking these paths would also require conditioning on \( D_1 \), which is not possible in a cross-sectional study. This is, however, done in longitudinal studies when researchers restrict participants to those without distress at study entry or adjust on baseline distress levels.

There are also other confounding variables which are fixed in time, at least for the study period. In Figure A1.2, these are represented simplistically by the single variable \( C \). Common examples taken into account in the job-stress literature include health status, marital status, socioeconomic level, and personality.

The above discussion focuses on estimating the association between workplace stressors and psychological distress in the current period (i.e. between \( W_2 \) and \( D_2 \)) from cross-sectional data. If longitudinal data are available, researchers are able to account for measurements in earlier periods 0 and 1 as well as in period 2. However, the modelling would need to account for the fact that, in this longitudinal setting, the private-life stressors are still confounders (e.g. \( W_2 \leftarrow D_1 \rightarrow L_2 \rightarrow D_2 \)) but also take on a role as mediators (\( W_1 \rightarrow D_1 \rightarrow L_2 \rightarrow D_2 \)). In this situation, alternative modelling techniques such as marginal structural models would be needed [4].

However, even in a cross-sectional study, adjusting on stressors in one domain to reduce confounding may introduce new bias owing to over-adjustment from removing some of the mediating effect of the stressor domain of interest [5]. Figure A1.3 introduces \( W_2^* \) to represent the variable we actually measure when asking about the true level of work pressure stressors with a questionnaire. Although meant to measure only work pressure stressors in time 2 (\( W_2 \)), this will rarely be the case since questionnaires commonly ask about stressors over a period going back in time. The measured work place stressor level \( W_2^* \) would
therefore capture portions of $W_1$ as well as $W_2$. Since $L_2$ has partly been caused by $W_1$ ($W_1 \rightarrow D_1 \rightarrow L_2$), conditioning on $L_2$ with the goal of reducing confounding of the $W_2 \rightarrow D_2$ estimate can introduce an over-adjustment bias. The net effect of this depends, of course, on the extent to which $W_2*$ picks up work pressure stressors occurring earlier and on the relative importance of the confounding and mediating pathways. The highlights the usefulness of presenting both unadjusted and adjusted estimates.

Obviously, the decision about which stressors to adjust on depends on the intended use of the analysis. From the perspective of a company’s occupational health department, adjusting on personal life stressors makes sense if the goal is, for example, to identify potential workplace causes of distress on which to intervene since an unconfounded estimate of the effect of workplace stressors on distress is needed. However, unadjusted measures of association with distress may be appropriate if the goal is to identify groups of employees at high risk of distress, for example to target interventions which may improve individuals’ abilities to cope with stressors in both their working and private lives. Here, allowing workplace stressor variables also to capture some of the distribution of the different life stressor may be desirable. Presenting both unadjusted and adjusted estimates may therefore be helpful for occupational health decision-makers.


Figure A1.1
Figure A1.2
Figure A1.3

Graph with nodes labeled $W_0$, $W_1$, $W_2$, $D_0$, $D_1$, $D_2$, $L_0$, $L_1$, $L_2$ and edges between them.
Approximate English translation of the questions used to measure workplace and private life stressors. All responses were Yes/No:

• During the last 6 months:
  o I am under constant time pressure because of an excessive work load.
  o I am frequently interrupted and bothered in my work.
  o I am expected to do an excessive amount of work.
  o I experience conflict situations at work.
  o I consider that I have too heavy responsibilities at work.
  o I consider myself to be in bad health.
  o I find my living environment too noisy.
  o I am dissatisfied with the place where I live.
  o I have financial difficulties.
  o I have difficulties at home with my spouse, my children, or another person close to me.

• Which of the following events have happened in your life in the last year:
  o Separation or divorce.
  o Moved house/apartment.
  o Accident, serious illness, or bereavement for a person close to me.
  o Physical aggression in the professional environment.
  o Physical aggression outside of the professional environment.
  o Mental/emotional (“psychique”) aggression in the professional environment or major upheaval of working conditions.
  o Mental/emotional (“psychique”) aggression outside of the professional environment.
  o More than a two weeks’ consecutive absence because of illness.
This appendix presents the methods and results of the models to predict the probability of psychological distress for a “typical” employee at different combinations of the work pressure, work conflict, and private life stressor domain scores. The motivation for this approach was to explore absolute risk (probability) of psychological distress as a complement to the multiplicative RR measure of association provided by the Poisson regression models. Absolute measures of association are of particular interest because they provide a more direct measure of the public health importance of a disease and allow an interpretation of model interaction terms in public health or biological terms [1,2].

To estimate the absolute risk of psychological distress, we developed logistic regression models for men and women separately on a training subset of the data, checked their predictive performance on the remaining test-set data, and then simulated the probability of distress from the model at fixed covariable values representing a “typical” company employee while varying the stressor domain scores. Each model started with the sociodemographic variables, the health concerns variable, and the work pressure, work conflict, and private life stressor scores plus all possible second-order interaction terms. We ran stepwise selection using the Akaike Information Criterion on a randomly selected set of 75% of the participants and then checked the “final” model’s discrimination performance with the c-statistic [3] and goodness-of-fit with the Hosmer-Lemeshow test [3] on the remaining 25% of participants.

We then drew 1000 simulations from the models’ coefficient distributions using the arm package in R [4]. These draws represent the uncertainty in the estimated coefficient distributions, providing a type of “informal Bayesian inference” since it approximates results.
from formal Bayesian models with non-informative prior distributions [5] (but is easier to calculate). For each coefficient draw, we calculated the probability of a GHQ score ≥3 for the covariate values representing a “typical” employee (youngest age group, married, not exclusively working days, lowest socioeconomic group, zero value for health concerns). We stratified in turn on each of the work conflict, work pressure, and life concerns scores whilst alternately varying one of the other two variables and holding the third at zero, allowing an exploration of the pattern of absolute risk of psychological distress under these different stressor combinations. The central 50% and 95% intervals of the predicted probabilities for each combination were plotted. A “typical” employee profile was used because of interaction terms involving the sociodemographic variables, meaning that the predicted probabilities differed for employees with different sociodemographic profiles. As we were interested in the pattern of absolute risk across different levels of the stressor scores, we fixed the other variables in the equation.

Figures A2.1 and A2.2 summarize the predicted probabilities of GHQ ≥3 under the different combinations of the work pressure, work conflict, and life concerns stressor scores. The fit statistics for each model were acceptable (footnotes, figures 1 and 2) and comparable with published predictive models [6,7]. The thick bars are 50% prediction intervals and the thin bars are 95% prediction intervals. The greater uncertainty (wider bars) for women is related to the smaller number of women in the study.

As noted in the main article, the graphs show that certain stressor combinations are associated with a particularly high risk of distress, e.g. women with high work pressure and work conflict scores. They also suggest that, among men, increasing values of the work conflict and private life stressor scores were associated with larger increases in absolute risk of distress than those
associated with increases in the work pressure score. Among women, different levels of the work conflict score were associated with the largest differences in absolute risk of distress. There was no evidence of effect modification on the additive scale for either men or women, since the distance between the predicted probabilities for the different stressor levels appeared similar across all levels of the other stressor scores. A possible exception was the work pressure score across the different work conflict strata for women. This, therefore, did not indicate the presence of individuals in the study population in whom the simultaneous presence of stressors in different domains was necessary for them to report psychological distress, i.e. the effect of stressors in one domain did not depend on the presence or absence of stressors in another domain [1,8]. It also did not flag any particular subgroup of workers who, assuming causal relationship between workplace stressors and distress, would disproportionately benefit more than others from a reduction in stressor exposures [1].

These analyses are obviously exploratory and, like all model-based results, depend on simplifying assumptions which need to be borne in mind. A key assumption is that the stressor domain scores are independent, while in reality a change in one would likely have amplification effects by acting through the others (appendix 1). Another assumption necessary if we wanted to give a causal interpretation to the associations is the absence of residual confounding. As discussed in the main article, this is claimed for our data.


Figure A2.1. Predicted probabilities of distress for male employee with typical characteristics

For each trio of predicted probability ranges, the left-most range is for the variable identified in the title as being varied set at a value of 0; the middle range is for this variable set at a value of 1; and the right-most range is for this variable set at \( \geq 2 \). The thick central line covers the 50% central interval and longer thin line covers the 95% central interval of the predicted probabilities. Fixed sociodemographic and other characteristics for typical employee are described in the methods.

a. After stepwise AIC-based selection, the retained model included marital status, shift work, work pressure, work conflict, health concerns, and the product terms of shift work and work pressure and of marital status and work conflict. Fit statistics: \( c = 0.74 \), Hosmer-Lemeshow \( p = 0.57 \).

b. Retained model included age, marital status, shift work, work pressure, life stressors, health concerns, and the product terms of work pressure and life stressors, of life stressors and health concerns, and of age and shift work. Fit statistics: \( c = 0.72 \), Hosmer-Lemeshow \( p = 0.60 \).

c. Retained model included age, marital status, socioeconomic status, shift work, work conflict, life stressors, health concerns, and the product terms of work conflict and life stressors, of age and shift work, of life stressors and health concerns, of socioeconomic status and life stressors, and of shift work and health concerns. Fit statistics: \( c = 0.77 \), Hosmer-Lemeshow \( p = 0.50 \).
Figure A2.2. Predicted probabilities of distress for female employee with typical characteristics

See the legend for Figure 1 for information on interpreting the graph.

a. After stepwise AIC-based selection, the retained model included marital status, work pressure, work conflict, and health concerns. Fit statistics: c = 0.72, Hosmer-Lemeshow p = 0.43.

b. Retained model included work pressure, life stressors, and health concerns. Fit statistics: c = 0.73, Hosmer-Lemeshow p = 0.12.

c. Retained model included socioeconomic status, shift work, work conflict, life stressors, health concerns, and the product terms of work conflict and life stressors and of socioeconomic status and health concerns. Fit statistics: c = 0.77, Hosmer-Lemeshow p = 0.21.