In the UK the most recent figures (20 observed and 16.35 expected cases) published by the Committee on Medical Aspects of Radiation in the Environment (COMARE)\(^4\) differ from those used by Körblein. Of note, COMARE reported a combined SIR of 1.07 (95% CI 0.92–1.26) from a random-effects meta-analysis of 37 estimates from five countries.\(^4\)

Körblein rightly points out that the 95% CI of the rate ratio estimate from the main analysis of childhood leukaemia in 0- to 4-year olds just includes the point estimate (2.19) of the German KiKK study. However, there is in fact considerable disagreement between the two studies. Assuming that the true IRR for leukaemia in 0- to 4-year olds comparing the 0- to 5-km zone with the 15-km zone around NPPs is indeed 2.2 (as estimated in the KIKK study), the probability of observing an IRR of 1.2 (as in our study) is 0.03 (Figure 1) and the power of rejecting the null hypothesis of no association at the 5% significance level is 76%.

Chirga argues that estimates of mean annual radiation doses originating from nuclear power plants (NPPs) are uncertain and that actual doses might fluctuate over time. He suggests that short spikes in emissions could increase the incidence of childhood cancer. Whereas we cannot exclude this possibility, we reiterate that our study provides little evidence that the rate of childhood cancer is higher in the proximity of NPPs. The few cases of cancer occurring in excess of the expected number of cases among children living in the 5-km zone around Swiss NPPs are well within statistical uncertainty, and so is the deficit of cases a few kilometres further afield. We agree with Chirga that our main exposure measure, Euclidean distance of place of residence to nearest NPP, was crude and can only serve as a proxy for true radiation exposure.

**Funding**

The CANUPIS study was supported by the Swiss Federal Office of Public Health (BAG 08.001616) and the Swiss Cancer League (KLS 02224-03-2008).

**Conflict of interest:** None declared.

**References**


Kivimaki et al. recently examined whether information on job strain improves risk prediction for coronary heart disease (CHD) beyond the standard Framingham risk score in a middle-aged low-risk working population.\(^1\) They observed that job strain was associated with an increased risk of CHD; however, when compared with the Framingham algorithm, adding job strain did not improve the model’s predictive performance.

The authors are to be commended for re-estimating (or refitting) the Framingham model instead of directly applying the model to their population. A prediction model tends to perform better in data from which it was derived than on a new dataset. This difference in performance is an indication of the optimism in the apparent performance in the derivation set.\(^2\) In light of the correlations between job strain and currently established risk factors, the effect size estimated for job strain may be a reflection of information lost during inappropriate modelling of information on the currently established Framingham CHD risk factors. When \(\beta\)-coefficients derived from the Framingham study population are directly applied while the coefficient for job strain is obtained from the study sample, then job strain has a so-called home-advantage. In other words, the effect size for...
job strain is likely to be over-optimized. A source of over-optimism relates to the choice of the existing benchmark methods applied for comparison purposes. Researchers are supposed to compare their new algorithm to state-of-the-art methods. Researchers may consciously or subconsciously choose suboptimal existing methods and exclude the best competing methods from the comparison. The new algorithm then seems better than competing approaches and over-optimistic results for the superiority of the new algorithm are reported. However, this is artificial as the best competing approaches have been disregarded. Since the definition of state-of-the-art methods is often ambiguous, such problems may occur even when researchers have decided to perform a fair comparison. I would like to refer to this mechanism (also known as the ‘straw-man phenomenon’), as ‘optimization of the competing methods’ following Jelizarow et al.

Some concerns regarding the null results obtained in the article are as follows.

(1) The authors unfortunately did not make it clear whether they used the most recently published ‘General Cardiovascular Risk Profile’ from the Framingham study or the previous General Cardiovascular Risk algorithm version of the ‘Framingham Risk Score’. By selecting the latter, they might have fallen into the ‘optimization of the competing methods’ trap, since the recent version of the Framingham prediction algorithm has been shown to be better calibrated—the continuous components have been log-transformed and the interaction between systolic blood pressure and use of blood pressure lowering agents has been allowed for.

(2) The authors have used the net reclassification improvement index (NRI) to examine whether adding information on job strain could improve the predictive performance of the Framingham algorithm. The NRI as first described by Pencina et al. conditions on case–control status. Since it depends on outcome status, it was not available for censored data, where not all persons will have follow-up completed until a certain time point, say 10 years. To overcome this, Cook and Ridker proposed selecting only persons with follow-up complete at a certain time point. They were applauded by Steyerberg and Pencina, who nonetheless proposed a simple alternative, based on the expected number of cases and controls, calculated using the Kaplan–Meier estimator. They did so because they considered that choosing 1 time point for analysis can lead to the exclusion of a (usually small) proportion of control participants and (a relatively large) proportion of cases, making the NRI estimates unstable. They appreciated that the asymptotic confidence intervals for NRIIs calculated by the approach outlined by Pencina et al. was no longer valid for the current extension and were consistent with Cook and Ridker and Pepe et al. who suggested using bootstrap resampling as a practical solution. Pencina and colleagues have recently extended NRI formulas to survival analysis. As such, the estimator no longer depends on the number, or even existence, of risk categories. It assumes that probabilities of events among those reclassified upwards or downwards can be obtained by pooling all individuals with the same reclassification. Cook and Paynter found this estimate so sensitive that it may detect differences that are negligible or unimportant. Net reclassification calibration methods are those preferred by Cook.

As Kivimaki et al. appreciate in their paper, failing to improve risk prediction would not (and should not) be translated as having no role in aetiology. The basic and commonly used measure of biological effect is the relative risk, expressed as a ratio of incidence rates, odds or hazards. However, recently, researchers have begun to believe that ‘it is timely for terms such as hypertension and hypercholesterolemia to be evacuated from our clinical vocabulary’ and that ‘the next generation of clinicians should treat risk not risk factors’. The relative risk, however, goes hand in glove with the ability to reclassify risk and, thus, should not be dismissed. It also makes sense to consider how intermediate risk categories change and whether this has been calculated correctly, because change will have the largest clinical impact in the intermediate risk group i.e. those in the ‘grey zone’ for whom the treatment decision is not clear.

References

6 Strom Moller C, Zethelius B, Sundstrom J, Lind L. Persistent ischaemic ECG abnormalities on repeated
ECG examination have important prognostic value for cardiovascular disease beyond established risk factors: a population-based study in middle-aged men with up to 32 years of follow-up. *Heart* 2007;93:1104–10.


Advance Access publication 27 January 2012
© The Author 2012; all rights reserved.

Authors’ response to: Can information on life stress improve CHD risk prediction in clinical practice?

From MIKA KIVIMÄKI,1* SOLJA T NYBERG,2,3 G DAVID BATTY,1 MARTIN J SHIPLEY,1 JANE E FERRIE,1 MARIANNA VIRTANEN,1,2 MICHAEL G MARMOT,1 JUSSI VAHTERA,2,4,5 ARCHANA SINGH-MANOUX6 and MARK HAMER1

1Department of Epidemiology and Public Health, UCL, London, UK, 2Centre of Expertise for Work Organizations, Finnish Institute of Occupational Health, Helsinki, Finland, 3Institute of Behavioral Sciences, University of Helsinki, Finland, 4Department of Public Health, University of Turku, Turku, Finland, 5Department of Primary Health Care, Turku University Hospital, Turku, Finland and 6Center for Research in Epidemiology and Population Health, INSERM, Villejuif, France

*Corresponding author. Department of Epidemiology and Public Health, University College London, 1-19 Torrington Place, London WC1E 6BT, UK. E-mail: m.kivimaki@ucl.ac.uk

We are pleased that our paper on job strain and coronary heart disease (CHD) risk prediction1 has attracted the interest of IJE readers. Given that stress has been the subject of observational research for several decades,2 surprisingly few attempts have been made to explore the predictive value of stress in clinical practice. The purpose of our paper was to address this gap in knowledge. We examined whether information on work stress would improve identification of patients at high risk of developing CHD if we already knew their standard risk factors, such as raised blood pressure, lipid levels, smoking habits and the presence of diabetes. Determining a patient’s risk of CHD is important as this helps a physician to decide whether there is a need to recommend lifestyle change or prescribe medication (e.g. statins). Our findings for job strain,3 the most commonly used operationalization of work stress, suggest this additional information is unlikely to enhance the predictive capacity of Framingham risk prediction algorithm (a summary of conventional risk factor levels).