


# BMJ Open Estimating expected life-years and risk factor associations with mortality in Finland: cohort study

Tommi Härkänen ,<sup>1</sup> Kari Kuulasmaa,<sup>1</sup> Laura Sares-Jäske,<sup>1</sup> Pekka Jousilahti,<sup>1</sup> Markku Peltonen,<sup>1</sup> Katja Borodulin,<sup>1,2</sup> Paul Knekt,<sup>1</sup> Seppo Koskinen<sup>1</sup>

**To cite:** Härkänen T, Kuulasmaa K, Sares-Jäske L, *et al.* Estimating expected life-years and risk factor associations with mortality in Finland: cohort study. *BMJ Open* 2020;**10**:e033741. doi:10.1136/bmjopen-2019-033741

► Prepublication history and additional material for this paper are available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2019-033741>).

Received 22 August 2019  
Revised 21 January 2020  
Accepted 24 January 2020



© Author(s) (or their employer(s)) 2020. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

<sup>1</sup>Public Health Solutions, Finnish Institute for Health and Welfare, Helsinki, Finland

<sup>2</sup>Age Institute, Helsinki, Finland

## Correspondence to

Dr Tommi Härkänen;  
tommi.harkanen@thl.fi

## ABSTRACT

**Objective** To develop a tool to inform individuals and general practitioners about benefits of lifestyle changes by providing estimates of the expected age of death (EAD) for different risk factor values, and for those who plan and decide on preventive activities and health services at population level, to calculate potential need for these.

**Design** Prospective cohort study to estimate EAD using a model with 27 established risk factors, categorised into four groups: (1) sociodemographic background and medical history, (2) lifestyles, (3) life satisfaction, and (4) biological risk factors. We apply a Poisson regression model on the survival data split into 1-year intervals.

**Participants** Total of 38 549 participants aged 25–74 years at baseline of the National FINRISK Study between 1987 and 2007.

**Primary outcome measures** Register-based comprehensive mortality data from 1987 to 2014 with an average follow-up time of 16 years and 4310 deaths.

**Results** Almost all risk factors included in the model were statistically significantly associated with death. The largest influence on the EAD appeared to be a current heavy smoker versus a never smoker as the EAD for a 30-year-old man decreased from 86.8 years, which corresponds to the reference values of the risk factors, to 80.2 years. Diabetes decreased EAD by >6.6 years. Whole or full milk consumers had 3.4 years lower EAD compared with those consuming skimmed milk. Physically inactive men had 2.4 years lower EAD than those with high activity. Men who found their life almost unbearable due to stress had 2.8 years lower EAD.

**Conclusions** The biological risk factors and lifestyles, and the factors connected with life satisfaction were clearly associated with EAD. Our model for estimating a person's EAD can be used to motivate lifestyle changes.

## INTRODUCTION

Most people want to live a long and healthy life. Choices affecting the prospects of achieving this goal are continually made by individuals themselves and by health professionals. Which amenable determinants of health and longevity should be emphasised in specific individual situations? It is well known from observational epidemiological studies that risk factors describing the sociodemographic background,<sup>1</sup> lifestyles,<sup>2–3</sup> dietary factors,<sup>4,5</sup> life satisfaction (LS)<sup>6–12</sup> and

## Strengths and limitations of this study

- Numerous risk factors of total mortality have been identified.
- Estimated ages of death are easier to interpret in comparing different risk factors and their levels than HRs.
- We compare differences in life expectancy for many different risk factors.
- Estimated differences in estimated ages of death cannot be interpreted as causal effects.

metabolic health<sup>13–15</sup> predict mortality. For example, vigorous physical activity has been found to decrease the risk of death by 22% compared with no physical activity. Smoking has been found to increase the hazard by 83% and life dissatisfaction by 49%.

Comparisons between different risk factors and their impact on survival could be carried out using expected age of death (EAD) that is easier to interpret than commonly presented HRs. Evidence-based decisions on how to improve the length of life, tailored to specific individual contexts, require reliable information on the EAD for different levels of these risk factors.

At different ages, the differences in EAD can vary considerably even if the HRs remain constant. The differences in life expectancies are generally larger for younger individuals, which illustrates the importance of lifestyle changes early in life. However, studies on the association between risk factors and EAD have rarely been reported in literature and they have generally only included a small number of risk factors simultaneously.

In this study, we analyse total mortality using a model with a large number of risk factors that have previously been found to be predictors of longevity. We include variables describing the socioeconomic background, medical history, lifestyle, LS and biological risk factors. We develop a multivariable

prediction model to estimate EAD and report the results also using HRs. As biological risk factors are influenced by lifestyle and socioeconomic background factors, we apply methods developed for graphical models to avoid potential overadjustment by these intermediate risk factors.

## METHODS

### Data

We used data of the National FINRISK Study where cross-sectional health examination surveys were conducted every 5 years from 1987 to 2007 in specific areas of Finland.<sup>16 17</sup> For each survey, a stratified, independent random sample was drawn from the general population using the national Population Information System. The age range was 25–64 years until 1992 and widened gradually in all areas to 25–74 years by 2007. The total sample size was 52 749 individuals, and the number of participants was 38 549 (73%). Participation rates decreased from 82% to 67% during these years. We excluded individuals who had more than 13 missing covariate values as we considered that the amount of information from these individuals was too small. We also excluded individuals who had total cholesterol below 3 or higher than 10, had body mass index (BMI) greater than 40, reported cancer or myocardial infarction diagnosed by a doctor at baseline, or disability or a disease which did not enable physical exercise 20–30 min as the risk of death of these individuals was likely to be too different from the other individuals. The analysis data set contained 35 804 individuals.

The cohorts formed by the survey participants were followed up for death using record linkage to the national Causes of Death Register maintained by Statistics Finland.<sup>18</sup> The follow-up for death ended on 31 December 2014. Mean follow-up time was 16.03 years. During the follow-up period, 4310 deaths occurred (2689 in men, 1621 in women).

### Potential risk factors

We categorised risk factors into four groups: sociodemographic background and medical history (abbreviated as background in the following), lifestyles, LS and biological risk factors. The first three groups were based on questionnaire data. Age, sex and education were included as sociodemographic background variables. Myocardial infarction of mother or father under age 60 (except that in the 2007 survey, the age limit for mothers was 65 years), and diabetes diagnosed with the respondent were included as medical background factors. The variables indicating lifestyles included dietary variables covering fresh vegetables and fruits, type of bread spread and type of milk, as well as indicators of smoking, alcohol use and physical activity.<sup>19 20</sup> LS variables comprised stress, accomplishments in life, stretching one's strength to the extreme at work, getting along with spouse or children, financial situation, having a friend and prospects of attaining the goals one would like to reach. Biological risk factors, BMI, systolic and diastolic blood pressure

and serum non-high-density lipoprotein (HDL) cholesterol were measured with high quality.<sup>16</sup> Details of these 27 risk factors are presented in the online supplementary material.

Age, number of cigarettes, alcohol use, BMI, blood pressure and cholesterol were modelled as continuous covariates, and all other risk factors as categorical covariates.

### Statistical methods

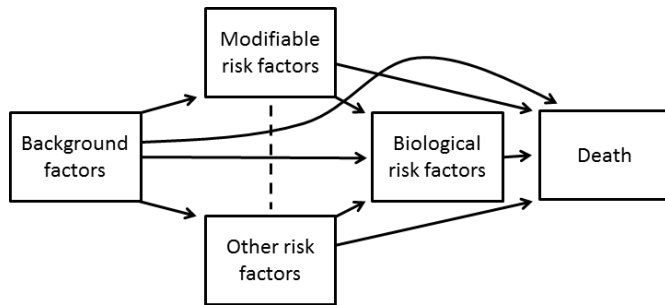
We performed multiple imputation (MI) to fill in the missing values.<sup>21</sup> Twenty-five imputed data sets were generated using the classification and regression trees (cart) method<sup>22</sup> by the mice package<sup>23</sup> of the R software.<sup>24</sup> As variables 'I feel it impossible to attain the goals that I'd like to reach' and 'I feel that I do not have any good friends' were asked only in two FINRISK surveys, these variables contained about 47% missing data, while the other variables related to LS contained about 21% missing data. Other variables contained 7% missing values or less. The descriptive statistics were calculated using the MI data.

We applied a Poisson regression model using the survreg function of the survival package. The survival data were split into 1-year intervals, within which the hazard function was assumed to be piecewise constant, using the Epi package.<sup>25</sup> The logarithm of the baseline hazard  $\lambda_0(t; \text{sex}_i)$  as a function of continuous age  $t$  in interaction with sex, and the logarithm of the HRs  $\beta^C(X_i^C)$  of the continuous risk factors  $X_i^C$  were modelled using natural cubic splines with 4 df, except the number of cigarettes and alcohol use, which were entered as linear terms in the linear predictor. The logarithms of the HRs  $\beta^N$  of the categorical covariates  $X_i^N$  were directly the regression coefficients in the Poisson regression model. Our hazard model was therefore

$$\lambda_i(t) = \lambda_0(t; \text{sex}_i) \exp \left\{ X_i^N \beta^N + \beta^C(X_i^C) \right\}.$$

In the following, we use the shorthand notations for the parameter vector  $\beta := \left( (\beta^H)', (\beta^N)', (\beta^C)' \right)'$  and risk factor vector  $X_i := (X_i^H, X_i^N, X_i^C)'$ , in which  $\beta^H$  and  $X_i^H$  correspond to the parameters, and age and sex related to the baseline hazard  $\lambda_0(t; \text{sex}_i)$ , respectively. In other words, the age-dependent baseline hazard was stratified with respect to sex, but other risk factors were assumed to act multiplicatively on the hazard.

We applied both simple and multiple regression models. First, the associations of each risk factor with death were estimated one by one, adjusting for age and sex. The second model was fully adjusted, thus we included all risk factors in the same model. All models contained the interaction term of calendar year before year 2000 and age below 75 to account for the decreasing hazard of death in the younger ages. In the projections we set this interaction term to correspond to the calendar year '2000 or later, or age 75 years or older', which is the reference category with HR equal to 1. In the online supplementary material, we present estimates also for the model in which



**Figure 1** Graphical model depicted by directed acyclic graph (DAG), which describes the assumed associations between the variables included in the analysis.

the biological risk factors were excluded but all other risk factors were included simultaneously.

The parameter estimates were pooled over the imputed data sets. The Wald tests were based on the pooled point estimates and covariance matrices.<sup>26</sup> We also applied the likelihood ratio test for the multiply imputed data to test if the risk factor interactions with sex were statistically significant in the model.<sup>27</sup>

The estimates of the Poisson regression model and the corresponding Cox proportional hazards (PH) model were compared, and virtually no differences were found (data not shown). We also tested the PH assumption of the Cox model, and the global p value<sup>28</sup> was 0.106 suggesting that the PH assumption was not violated.

The projected EAD value for an individual was estimated by calculating the linear predictor  $LP_{it} = \log(\lambda_i(t)) = X_{it}\beta$  values using the parameter estimates of the Poisson regression model and risk factor values for each age year starting from the baseline age  $a$  of the individual. The conditional probability  $P(T \in (t, t+1) | T > t)$  for each year  $(t, t+1)$  of death given survival until that year was  $1 - \exp\{-\exp(LP_{it})\} = 1 - \exp\left\{-\int_t^{t+1} \lambda_i(s) ds\right\}$ . Note that in the following we use integer valued  $t$  so that the integral in the previous expression simplifies to  $\int_t^{t+1} \lambda_i(s) ds = \lambda_i(t)$ . Finally, the EAD was calculated using the basic definition of expectation of a discrete probability distribution as

$$\begin{aligned} \text{EAD} &= E[T | \lambda_0, X_i, \beta, T \geq a] = \sum_{t=a}^{110} t \times P(T \in (t, t+1)) \\ &= \sum_{t=a}^{110} t \times P(T \in (t, t+1) | T > t) P(T > t) \end{aligned}$$

In the last expression, the two terms correspond to the conditional probability and the survival function  $S(t) = \exp\left\{-\int_a^t \lambda_i(s) ds\right\} = \prod_{k=a}^{t-1} \exp\left\{-\int_k^{k+1} \lambda_i(s) ds\right\} = \prod_{k=a}^{t-1} \exp\{-\exp(LP_{ik})\}$ , respectively. Note that standard cohort or period methods to estimate the EAD are not sufficient, as we need a parametric survival model to account for the large number of risk factors. The calculation of the EAD using R is illustrated in the online supplementary material.

The CIs for the EADs and their contrasts were calculated by parametric bootstrap with 1000 samples drawn from the multinormal distribution defined by the parameter estimates and their variance estimates, and then

calculating the EAD estimate for each sample and using the 2.5% and 97.5% quantile points as the CI limits.

As there are various causal pathways between the risk factors, we applied the causal calculus by Judea Pearl<sup>29</sup> for estimating EADs. The possible dependencies between the risk factors are illustrated in figure 1. We assumed that the biological risk factors can depend on the background, modifiable and other risk factors. In order to simplify the modelling assumptions, we assumed that the modifiable and other risk factors depended only on the background factors, and the additional dependencies between the modifiable and other risk factors were modelled only as associations without assuming (temporal) ordering between them.

Application of the causal effect package<sup>30</sup> of the R software<sup>24</sup> provided us with the formula to calculate the distribution of the time of death conditionally on the background, modifiable and other risk factors, thus the mediating biological risk factor variables were handled by integrating them out. This was conducted by generating predictive values for the biological risk factors using the numerical Monte Carlo method and MI based on the random forest method<sup>31</sup> in the mice package,<sup>23</sup> and then averaging the EADs based on the 1000 imputed data sets. The variance of these predictive distributions is large, thus the prediction intervals of these causal effect estimates are considerably larger than the full conditional prediction intervals based on fixing the values of all risk factors including the biological risk factors.

### Patient and public involvement

This research was done without participant involvement. The participants of the FINRISK Study were not invited to comment on the study design and were not consulted to interpret the results. Participants were not invited to contribute to the writing or editing of this document for readability or accuracy.

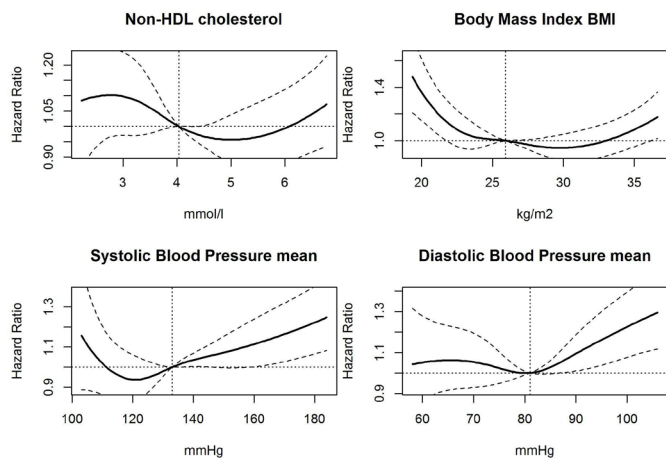
### RESULTS

Descriptive statistics of the baseline measurements of the participants aged 25–74 years are shown in online supplementary table S1.

For the associations between risk factors and the hazard of death, likelihood ratio test ( $p=0.44$ ) indicated that there was no need to include interaction terms of sex and risk factors, except age. Therefore, only the interaction of sex and age was included, and other risk factors were entered only as main effects in the models.

Almost all risk factors were statistically significantly associated with the hazard of death even when adjusted for all other background, lifestyle, LS and biological risk factors (table 1a, table 1b, table 1c, figure 2 and online supplementary figure S1). Level of education remained a significant predictor even after adjustment for all the other risk factors (table 1a). The higher the education, the lower the hazard (HR=0.90 for highest vs lowest education). Mother's myocardial infarction before age





**Figure 2** HRs of continuous risk factors based on the fully adjusted model. Solid curves correspond to the point estimates of the HR contrasted with the hazard at the median value of the risk factor denoted by the vertical dotted line. Dashed lines correspond to the 95% CIs. HDL, high-density lipoprotein.

60 was associated with higher hazard (HR=1.20), whereas the corresponding association between the diagnosis of the father and hazard was not significant. Diagnosed diabetes for which treatment by medicine was prescribed increased the hazard (HR=2.21).

Current smokers had higher hazard compared with never smokers (HR=1.67), and every 10 cigarettes per day increased the hazard (HR=1.18, [table 1b](#)). If alcohol consumption exceeded 84 g/week, an additional 100g/week increased the hazard (HR=1.08), and if one reported feeling intoxicated less than once a month, the hazard decreased (HR=0.81) compared with those feeling intoxicated more frequently than once a month.

Whole or full milk increased the hazard (HR=1.53) compared with drinking skimmed milk ([table 1b](#)). Use of butter, butter-vegetable oil mixture or cooking margarine also increased the hazard (HR=1.14). Daily use of fresh vegetables decreased the hazard (HR=0.88) compared with using them twice a week or less frequently. Also, eating fruits and berries daily or almost daily decreased the hazard (HR=0.85) compared with consuming them twice a month or less frequently.

Engagement in high (HR 0.75) or medium (HR 0.81) volume of leisure time physical activity was associated with a lower hazard compared with those with no leisure time physical activity ([table 1b](#)).

Most of the variables indicating satisfaction with life were also significant predictors ([table 1c](#)). Having some stress, but no more than what is usual was associated with lower hazard (HR=0.73) than when feeling life almost unbearable. Hazard seemed to be higher for those who did not work (HR=1.13), but no clear trend can be seen among the other categories of the work stress variable.

The less one was satisfied with his/her accomplishments in life, the higher was the hazard (HR=1.32 for unsatisfied, [table 1c](#)). Similarly, the more one disagreed with feeling impossible to obtain one's goals, the lower the

hazard (HR=0.73 for somewhat disagree). If one's financial status was somewhat better than before (HR=0.92), the hazard was lower than if the status was much better, about the same or worse.

Not having a spouse was associated with a higher hazard (HR=1.35) compared with often having trouble with the spouse ([table 1c](#)). Having never special trouble with children was associated with lower hazard (HR=0.89) than not having children or having often special trouble with children.

For non-HDL cholesterol the hazard was lowest at around 5 mmol/L, for BMI at slightly below 30 kg/m<sup>2</sup>, for systolic blood pressure at 120 mm Hg and for diastolic blood pressure at slightly above 80 mm Hg ([figure 2](#)).

The HR estimates were very similar with and without the biological risk factors (online supplementary table S2).

The largest influence on the EAD appeared to be a current smoker versus a never smoker as the EAD for a 30-year-old man decreased from 86.8 years, which corresponds to the reference values of the risk factors, to 82.6 years ([table 2](#) contains the most influential risk factors reducing the EAD of a 30-year-old man by more than 2 years, online supplementary table S3 all risk factors and online supplementary table S4 the contrasts of the risk factor categories), and additionally, smoking 20 cigarettes per day decreased EAD further to 80.2 years while keeping all other risk factors at the same values. Diabetes decreased EAD almost as much to 80.3 years. Whole or full milk consumers had EAD of 84.5 years compared with 87.9 years of those consuming skimmed milk. Physically inactive men had EAD of 85.0 years whereas those with high activity had EAD of 87.4 years. Men, who found their life almost unbearable due to stress, had EAD of 84.0 years. For older men and for women the differences were similar but smaller. The estimates based on causal calculus were lower than the full conditional EADs based on fixing the biological risk factors in the mode values, which are relatively close to the low risk values ([figure 2](#)). BMI values below 22 and above 33, non-HDL cholesterol values below 3.6 and above 6.5, diastolic blood pressure above 85 and systolic blood pressure values below 110 and above 135 appeared to reduce the EAD when compared with the lowest risk values (online supplementary figure S2), but these optimal values are based on the other risk factors being at their optimal values. In practice, for example, overweight and obesity can increase blood pressure compared with normal weight, which can increase mortality.

## DISCUSSION

The biological risk factors and lifestyles, and the factors connected with LS were clearly associated with the EAD, and these associations did not disappear when adjusted for a large number of risk factors. Factors like frequent smoking, certain dietary choices with saturated fat, low leisure time physical activity, having been diagnosed with diabetes and experiencing stress were associated with lower years of EAD across all age groups and in both

**Table 1a** Estimates of HRs (with 95% CIs) of background risk factors based on age and sex-adjusted, and fully adjusted regression models, and the Wald test p values

	Deaths	Age and sex-adjusted regression model*		Multiple regression model†	
Education				P=0.032	
Low	1524	1.00		1.00	
Middle	1421	0.86	(0.78 to 0.95)	0.94	(0.86 to 1.03)
High	1365	0.70	(0.64 to 0.77)	0.90	(0.81 to 0.99)
Has your father been diagnosed as having had myocardial infarction when he was under 60 years?				P=0.082	
No	3516	1.00		1.00	
Yes	794	1.13	(1.02 to 1.25)	1.10	(0.99 to 1.22)
Has your mother been diagnosed as having had myocardial infarction when he was under 60 years?				P=0.013	
No	3918	1.00		1.00	
Yes	392	1.28	(1.11 to 1.48)	1.20	(1.04 to 1.39)
Have you ever been diagnosed as having diabetes, and how is it treated?				P<0.001	
No	3907	1.00		1.00	
Yes, dietary counselling	184	1.19	(0.98 to 1.44)	1.16	(0.96 to 1.41)
Yes, medicine prescribed	220	2.24	(1.92 to 2.63)	2.21	(1.87 to 2.60)

\*Age and sex-adjusted model.

†All risk factors in [table 1a](#), [b](#) and [c](#), including also age, sex and the biological risk factors in figure 2, were entered in the model simultaneously.

**Table 1b** Estimates of HRs (with 95% CIs) of lifestyle risk factors based on age and sex-adjusted, and fully adjusted regression models, and the Wald test p values

	Deaths	Age and sex-adjusted regression model*		Multiple regression model†	
Smoking?				P<0.001	
Never	1840	1.00		1.00	
Quitted more than 6 months ago	964	1.22	(1.09 to 1.35)	1.17	(1.05 to 1.31)
Current	1506	2.75	(2.50 to 3.01)	1.67	(1.44 to 1.93)
Number of cigarettes per day				P<0.001	
	1499	1.04	(1.04 to 1.05)	1.02	(1.01 to 1.02)
Alcohol per week (per 100g above 84g)				P<0.001	
	1056	1.19	(1.16 to 1.22)	1.08	(1.04 to 1.12)
How often, during the past 12 months, have you felt intoxicated by alcohol?				P=0.001	
Once a week or more often	393	1.00		1.00	
At least once a month	752	0.78	(0.67 to 0.91)	0.95	(0.81 to 1.12)
Less than once a month	3165	0.51	(0.44 to 0.58)	0.81	(0.69 to 0.94)
Bread spread				P=0.004	
Low-fat spread, nothing, household margarine or spreads with plant stanols	2494	1.00		1.00	
Butter, butter-vegetable oil mixture or cooking margarine	1816	1.32	(1.22 to 1.43)	1.14	(1.04 to 1.24)
If you drink milk, what kind is it usually?				P<0.001	
Skimmed milk	499	1.00		1.00	
Low-fat milk or I do not drink milk	2705	1.25	(1.12 to 1.41)	1.17	(1.04 to 1.31)
Whole or full milk	1106	2.10	(1.84 to 2.39)	1.53	(1.33 to 1.77)
How often do you usually eat fresh vegetables, root crops (excluding potatoes) and fresh salads?				P=0.031	
At most twice a week	2245	1.00		1.00	
Almost daily	1277	0.72	(0.65 to 0.78)	0.94	(0.86 to 1.04)
Daily	788	0.61	(0.55 to 0.68)	0.88	(0.79 to 0.99)

Continued

Table 1b Continued

	Deaths	Age and sex-adjusted regression model*		Multiple regression model†	
How often do you usually eat fruits or berries?				P=0.007	
At most twice a month	462	1.00		1.00	
Once or twice a week	1285	0.67	(0.59 to 0.76)	0.94	(0.82 to 1.08)
Almost daily or daily	2562	0.48	(0.42 to 0.54)	0.85	(0.74 to 0.98)
How much do you exercise and stress yourself physically in your leisure time?				P<0.001	
Inactive	1241	1.00		1.00	
Low	2553	0.66	(0.60 to 0.72)	0.81	(0.74 to 0.89)
High	516	0.47	(0.41 to 0.54)	0.75	(0.65 to 0.86)

\*Age and sex-adjusted model.

†All risk factors in table 1a, b and c, including also age, sex and the biological risk factors in figure 2, were entered in the model simultaneously.

Table 1c Estimates of HRs (with 95% CIs) of life satisfaction risk factors based on age and sex-adjusted, and fully adjusted regression models, and the Wald test p values

	Deaths	Age and sex-adjusted regression model*		Multiple regression model†	
Have you felt tensioned, stressed or under a lot of strain during the past month?				P=0.002	
Yes, my life is almost unbearable	104	1.00		1.00	
Yes, quite more so than people usually are	521	0.60	(0.47 to 0.76)	0.82	(0.64 to 1.06)
Yes, somewhat, but no more than what is usual	2288	0.47	(0.37 to 0.59)	0.73	(0.57 to 0.93)
Not at all	1397	0.50	(0.39 to 0.63)	0.79	(0.61 to 1.02)
How satisfied are you with your accomplishments in life?				P=0.004	
Very satisfied	282	1.00		1.00	
Satisfied	1289	1.16	(0.98 to 1.37)	1.12	(0.94 to 1.33)
Somewhat satisfied	2325	1.49	(1.27 to 1.75)	1.21	(1.02 to 1.43)
Unsatisfied	414	2.41	(1.98 to 2.93)	1.32	(1.06 to 1.64)
How often do you have to stretch your strength to the extreme in order to cope with your present work?				P=0.001	
I do not work	1603	0.96	(0.74 to 1.24)	1.13	(0.86 to 1.48)
Almost all the time	115	1.00		1.00	
Quite often	376	0.75	(0.56 to 1.00)	0.92	(0.68 to 1.24)
Sometimes	780	0.74	(0.57 to 0.97)	0.99	(0.75 to 1.31)
Seldom or never	1436	0.70	(0.54 to 0.91)	0.96	(0.73 to 1.26)
Is your financial status now better or worse than before?				P=0.001	
Much better	448	1.00		1.00	
Somewhat better	1024	0.98	(0.83 to 1.15)	0.93	(0.79 to 1.09)
About the same	1875	1.20	(1.04 to 1.39)	1.09	(0.94 to 1.26)
A bit worse	654	1.37	(1.16 to 1.62)	1.12	(0.94 to 1.33)
A lot worse	309	1.87	(1.55 to 2.25)	1.14	(0.93 to 1.39)
I feel it impossible to attain the goals that I'd like to attain.				P<0.001	
Absolutely agree	263	1.00		1.00	
Somewhat agree	775	0.71	(0.58 to 0.86)	0.87	(0.71 to 1.07)
Hard to say	1981	0.69	(0.58 to 0.83)	0.85	(0.70 to 1.03)
Somewhat disagree	827	0.49	(0.41 to 0.60)	0.72	(0.59 to 0.89)
Absolutely disagree	465	0.46	(0.37 to 0.57)	0.74	(0.59 to 0.94)
I feel that I do not have any good friends.				P=0.105	
Disagree	3432	1.00		1.00	

Continued

Table 1c Continued

	Deaths	Age and sex-adjusted regression model*		Multiple regression model†	
Agree or hard to say	878	1.38	(1.24 to 1.53)	1.09	(0.98 to 1.22)
Do you have trouble getting along with your spouse?				P<0.001	
I do not have a spouse	1276	1.60	(1.26 to 2.03)	1.35	(1.06 to 1.73)
Quite often or more often	138	1.00		1.00	
Sometimes or seldom	2175	0.98	(0.78 to 1.25)	1.14	(0.90 to 1.45)
Never	721	1.02	(0.79 to 1.31)	1.21	(0.94 to 1.56)
How often have your children caused you special trouble?				P=0.040	
Often or I have no children	1021	1.00		1.00	
Sometimes	511	0.78	(0.67 to 0.90)	0.96	(0.83 to 1.12)
Seldom	1100	0.72	(0.64 to 0.81)	0.93	(0.82 to 1.06)
Never	1678	0.71	(0.64 to 0.79)	0.89	(0.80 to 1.00)

\*Age and sex-adjusted model.

†All risk factors in table 1a, b and c, including also age, sex and the biological risk factors in figure 2, were entered in the model simultaneously.

sexes. We found that adjustment for a wide range of established risk factors had only marginal effect on the hazard rate estimates when compared with age-adjusted estimates in analyses based on large national cohort data. This suggests that there is a wide variety of potential interventions, which could increase the length of life among people with low life expectancies, as they are likely to have suboptimal values in several risk factors. The effects of interventions on EAD are likely to be largest in younger ages as for the oldest individuals the hazard of death is inevitably high. The causal calculus had only marginal influence on the EAD estimates when compared with the full conditional estimates also indicating that the projections are not sensitive to the overadjustment.

### Comparison with other studies

Most of our findings for significant factors that were associated with mortality have been reported previously. However, calculations for life expectancies across risk factor categories have not been previously carried out in a large data set and with a large number of risk factors. There are several articles in which the number of risk factors was smaller, the number of modifiable lifestyle factors was small or the methods are more difficult to implement on individual level.<sup>32–34</sup> Thus, we bring new information on how EAD may change across risk factor categories. Presenting the results on EAD makes the interpretation of the findings more comprehensible than using traditional HRs.

The largest reductions in EAD were found among smokers and diabetics, more than 6 years for a 30-year-old man when compared with never smokers and non-diabetics. Smoking has been found to increase all-cause mortality 1.83-fold in older people.<sup>3</sup> New forms of tobacco use, especially e-cigarettes, have become popular during the past few years, but their long-term associations

with mortality are currently unknown.<sup>35</sup> E-cigarettes have been found to reduce the exposure to carcinogens and toxicants compared with traditional cigarettes,<sup>36</sup> but they may increase the risk of respiratory diseases compared with never smokers.<sup>37</sup> Type 2 diabetes has been found to increase all-cause mortality 1.85-fold.<sup>38</sup>

As many other observational cohort studies, also the FINRISK Study shows a J-shaped association with alcohol consumption and mortality, with moderate drinking associating to lower hazard than non-drinking. This has been the case in particular for cardiovascular diseases. However, there is ongoing debate as to how much of this is causal and how much is explained by possible reversed causality. A recent study based on Mendelian randomisation found no J shape in vascular disease incidence.<sup>20</sup>

Results from this study support previous finding also on the inverse associations of vegetables, and fruits and berries with mortality.<sup>5</sup> In the present study, consumption of milk or bread spread with high content of saturated fat was directly associated with the hazard. This is in line with dietary guidelines suggesting limiting intake of saturated fat. However, most of the previous observational studies have failed to find an association between intake of saturated fat and mortality, although this may be partly due to methodological limitations.<sup>4</sup>

Leisure time physical activity is known to associate with mortality being in line with our findings.<sup>39</sup>

We found stress to increase all-cause mortality, but effect modifiers such as income have been found in the literature.<sup>12</sup> LS has been found to be associated with lower hazard<sup>10</sup> also in an earlier Finnish study where the adjusted HR was 1.49 for dissatisfied compared with satisfied men, but among women a corresponding association between LS and survival was not found. In Germany, a similar interaction with sex was found,<sup>8</sup> and relevant

**Table 2** Projections of the expected age of death for individuals in different ages and sex. Risk factors are at their reference values, and the value of one risk factor was changed at a time

Modified risk factor*†‡	Men			Women		
	30	50	70	30	50	70
Age						
All risk factors were fixed to their reference levels§	88.1	88.6	90.1	89.3	89.5	90.4
Causal effect fixing risk factors except the biological risk factors to their mode values	86.8	87.3	89.0	88.4	88.6	89.6
Smoking?						
Never*	86.8	87.3	89.0	88.4	88.6	89.6
Quitted more than 6 months ago	85.6	86.2	88.1	87.4	87.7	88.9
Current, 0 cigarettes per day	82.6	83.4	85.8	85.0	85.4	86.9
Current, 20 cigarettes per day	80.2	81.2	84.1	82.9	83.4	85.3
Have you ever been diagnosed as having diabetes, and how is it treated?						
No*	86.8	87.3	89.0	88.4	88.6	89.6
Yes, dietary counselling	85.4	86.0	88.0	87.3	87.6	88.8
Yes, medicine prescribed	80.2	81.2	84.2	83.1	83.6	85.5
If you drink milk, what kind is it usually?						
Skimmed milk	87.9	88.4	89.9	89.4	89.6	90.5
Low-fat milk or I do not drink milk*	86.8	87.3	89.0	88.4	88.6	89.6
Whole or full milk	84.5	85.2	87.3	86.5	86.8	88.1
How much do you exercise and stress yourself physically in your leisure time?						
Inactive	85.0	85.6	87.6	87.0	87.3	88.5
Low*	86.8	87.3	89.0	88.4	88.6	89.6
High	87.4	87.9	89.5	88.8	89.0	90.0
Have you felt tensioned, stressed or under a lot of strain during the past month?						
Yes, my life is almost unbearable	84.0	84.7	86.9	86.1	86.5	87.8
Yes, quite more so than people usually are	85.6	86.2	88.1	87.4	87.7	88.9
Yes, somewhat, but no more than what is usual*	86.8	87.3	89.0	88.4	88.6	89.6
Not at all	86.2	86.8	88.6	87.9	88.1	89.2

\*Mode values of risk factors were used as the risk factor values for the reference individual.

†Intermediating biological risk factors were handled using the causal calculus.

‡Projections for other risk factors are found in online supplementary table S3.

§For full conditional projections also fixed body mass index (BMI)=24.4, non-high-density lipoprotein (HDL) cholesterol=3.8, diastolic blood pressure=79, systolic blood pressure=130.

determinants of LS were found to be psychological, social and lifestyle factors and perceived health.

The BMI values associated with the lowest hazard of death were around 30, and the highest hazard was at BMI values around 20 and above 35. In a simple regression analysis we found the lowest hazard around BMI value of 25 (data not shown), as higher BMI values are often associated with various other risk factors, which we adjusted for in the fully adjusted model. This difference can be partly explained by the causal path from obesity to death via hypertension, as the simple regression analysis is not adjusted for these intermediating variables. This phenomenon illustrates the need to change the BMI value and other risk factor values, which are associated (possibly causally) with the changes in the BMI, in order to obtain EAD changes, which could take place after a change in

BMI (or other risk factors). For example, increase in BMI from 25 to 30 in the age group of 50–60 years is associated with 4.6 mm Hg increase in the systolic and 3.9 mm Hg increase in diastolic blood pressure on average, which eliminate the 5% lower hazard seen in the HR estimate of BMI (figure 2). These results align with those of a Mendelian randomisation study,<sup>40</sup> in which the unconfounded estimate of the lowest hazard was found to be between BMI values of 22 and 25. The increased risk from low BMI values may also be driven by reversed causality from sickness causing weight loss.

We tested the interactions of the risk factors with sex in a model with age adjustment. Some of the interactions were found statistically significant (data not shown). However, in our large data set, some of these significant associations were of small magnitude. A particularly



strong interaction was observed between sex and the variable on having trouble getting along with the spouse: single men appeared to have a higher hazard than other men, but for single women the hazard was close to that in other women. Feeling it impossible to attain the goals that one would like to reach appeared to increase the hazard slightly more for men than for women.

Despite the large number of risk factors being simultaneously adjusted for in our analysis, we found that low education was still associated with a higher mortality, which has been found also in numerous studies in different countries.<sup>1 41 42</sup>

Our results are in concordance with general health promotion guidelines. The important message of our results is that the HR estimates based on the age and sex-adjusted, and fully adjusted regression models were close to each other. A difference in a lifestyle factor is generally associated with differences in several other risk factors. For example, an increase in physical activity can reduce weight, blood pressure and cholesterol levels, which are all associated with a lower hazard of death. Therefore, the EAD differences might be even larger than our results, which were based on considering the differences of a single risk factor at a time.

### Strengths and limitations

The strengths of this study include the use of a large cohort with highly standardised baseline measurements and a long follow-up period. Participation rates were high from 1987 to 1997, thus most deaths were from surveys where non-participation was unlikely to cause serious selection bias in our estimates. The risk factors appeared to satisfy the PH assumption, thus the extrapolation of the results in the oldest ages and also in the future seems realistic, but future work on interactions with calendar time could provide further insight to the associations of the risk factors and mortality. We have modelled the association of the continuous risk factors, including age, with death using the spline functions as in many cases the associations are not linear and a categorisation of continuous risk factors would result in information loss. Generally, latent confounders can compromise the results of observational studies, but we used a large number of risk factors in our models to mitigate such bias. The application of Judea Pearl's causal calculus<sup>29</sup> should mitigate the potential overadjustment by the intermediating biological risk factors (BMI, cholesterol and blood pressure).

The main weakness of this study concerns the possible causal interpretations of our results. Differences between the projected EADs represent differences between population subgroups, but not necessarily the effects of changing an individual's risk factor values. For example, if an individual increases physical activity, it is likely that, for example, BMI, cholesterol and blood pressure also change. The scientific evidence on the effects of risk factor changes varies between risk factors. For example, quitting of smoking can reduce the risk of coronary heart disease quickly, but the risk of cancer decreases slowly over 10–15 years, but for several

other risk factors evidence on causal relationships is vague. The accumulated risk factor history can have a considerable effect on the EAD via increased hazard of cancer and increased hazard of death after the cancer diagnosis. Therefore, a more appropriate interpretation of the differences is a comparison of two population groups. We feel that none of these risk factors could be potentially harmful, if a user attempts to change his/her risk factor value into a more optimal category. The effects of medication were not accounted for in our analyses, thus the associations of high blood pressure or cholesterol levels with mortality can be underestimated as more effective medication, which could have started after the baseline measurement, could have reduced the hazard during the follow-up. Risk factor values could have changed during the follow-up, for example, many smokers quit smoking at some time, thus the HRs of current smokers can also be underestimated. Our proxy variables for saturated fat in food were limited to type of milk and bread spread, thus omitting other sources of saturated fat. Type of milk and bread spread, however, are two of the main sources of saturated fat.<sup>4</sup> Different risk factors are likely to contain different amount of measurement error. The most accurately measured risk factors were BMI, blood pressure and laboratory measurements, but self-reported lifestyle and LS factors are likely to vary considerably both between individuals and over time. We did not incorporate interactions in our models, except with age and sex, although some interactions have been found in the literature. As our data set contained only individuals below 75 years of age, our projections do not account for possible interactions of age and the other risk factors in the HRs in the oldest age groups. If the (continuous) risk factors have extreme values for some individuals, then our projections might not be very reliable as our data set contained only a small number of such individuals.

### CONCLUSIONS AND POLICY IMPLICATIONS

EAD is an easy to understand measure for comparing survival associated with different risk factor values. The biological risk factors and lifestyles, and the factors connected with LS were clearly associated with EAD. Our model for estimating a person's EAD can be used to motivate lifestyle changes. Furthermore, decision makers, who might have a possibility to influence working and other relevant environments, can use it to estimate potential need for preventive measures and medical care in the population.<sup>43</sup>

**Twitter** Markku Peltonen @MarkkuPeltonen

**Acknowledgements** This work was supported by the Academy of Finland under grant numbers 266251 and 307907, and by the Duodecim Medical Publications.

**Contributors** TH, KK and SK had central role in planning, conduct and reporting of the work, accept full responsibility for the work and/or the conduct of the study, had access to the data and controlled the decision to publish. LSJ, PJ, MP, KB and PK participated in planning, conduct and reporting of the work.

**Funding** The study received external funding from Duodecim Medical Publications and was supported by the Academy of Finland (grant numbers 266251 and 307907).

**Disclaimer** Researchers were independent of the funders. All authors had full access to all of the data (including statistical reports and tables) in the study and can take responsibility for the integrity of the data and the accuracy of the data analysis.

**Competing interests** TH reports grants from Duodecim Medical Publications and grants from Academy of Finland during the conduct of the study.

**Patient consent for publication** Not required.

**Ethics approval** Ethical approval has been obtained according to the commonly required research procedures and Finnish legislation during each survey. The last three surveys were approved by the coordinating ethics committee of the Helsinki and Uusimaa Hospital District.

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Data availability statement** No data are available. The individual-level data cannot be distributed as they are sensitive data. However, there is a procedure for requesting access to individual data for research collaboration on <https://thl.fi/en/web/thlfi-en/research-and-expertwork/population-studies/the-national-finrisk-study>.

**Open access** This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

#### ORCID iD

Tommi Härkänen <http://orcid.org/0000-0002-4577-1808>

#### REFERENCES

- Meara ER, Richards S, Cutler DM. The gap gets bigger: changes in mortality and life expectancy, by education, 1981–2000. *Health Aff* 2008;27:350–60.
- Paffenbarger RS, Hyde RT, Wing AL, et al. The association of changes in physical-activity level and other lifestyle characteristics with mortality among men. *N Engl J Med Overseas Ed* 1993;328:538–45.
- Gellert C, Schöttker B, Brenner H. Smoking and all-cause mortality in older people. *Arch Intern Med* 2012;172:837–44.
- de Souza RJ, Mente A, Maroleanu A, et al. Intake of saturated and trans unsaturated fatty acids and risk of all cause mortality, cardiovascular disease, and type 2 diabetes: systematic review and meta-analysis of observational studies. *BMJ* 2015;351:h3978.
- Schwingshackl L, Schwedhelm C, Hoffmann G, et al. Food groups and risk of all-cause mortality: a systematic review and meta-analysis of prospective studies. *Am J Clin Nutr* 2017;105:1462–73.
- Boehm JK, Winning A, Segerstrom S, et al. Variability Modifies Life Satisfaction's Association With Mortality Risk in Older Adults. *Psychol Sci* 2015;26:1063–70.
- Margolis R, Myrskylä M. Family, money, and health: regional differences in the determinants of life satisfaction over the life course. *Adv Life Course Res* 2013;18:115–26.
- Lacruz ME, Emeny RT, Baumert J, et al. Prospective association between self-reported life satisfaction and mortality: results from the MONICA/KORA Augsburg S3 survey cohort study. *BMC Public Health* 2011;11:2458–579.
- Collins AL, Gleit DA, Goldman N. The role of life satisfaction and depressive symptoms in all-cause mortality. *Psychol Aging* 2009;24:696–702.
- Koivumaa-Honkanen H, Honkanen R, Viinamäki H, et al. Self-Reported life satisfaction and 20-year mortality in healthy Finnish adults. *Am J Epidemiol* 2000;152:983–91.
- Holt-Lunstad J, Smith TB, Layton JB. Social relationships and mortality risk: a meta-analytic review. *PLoS Med* 2010;7:e1000316.
- Redmond N, Richman J, Gamboa CM, et al. Perceived stress is associated with incident coronary heart disease and All-Cause mortality in Low- but not High-Income participants in the reasons for geographic and racial differences in stroke study. *J Am Heart Assoc* 2013;2:e000447.
- Flegal KM, Kit BK, Orpana H, et al. Association of all-cause mortality with overweight and obesity using standard body mass index categories: a systematic review and meta-analysis. *JAMA* 2013;309:71–82.
- Ford ES. Risks for all-cause mortality, cardiovascular disease, and diabetes associated with the metabolic syndrome: a summary of the evidence. *Diabetes Care* 2005;28:1769–78.
- Lakka H-M, Laaksonen DE, Lakka TA, et al. The metabolic syndrome and total and cardiovascular disease mortality in middle-aged men. *JAMA* 2002;288:2709–16.
- Borodulin K, Tolonen H, Jousilahti P, et al. Cohort profile: the National FINRISK study. *Int J Epidemiol* 2017.
- Jousilahti P, Laatikainen T, Peltonen M, et al. Primary prevention and risk factor reduction in coronary heart disease mortality among working aged men and women in eastern Finland over 40 years: population based observational study. *BMJ* 2016;352:i721.
- Deaths. Available: [http://www.stat.fi/til/kuol/meta\\_en.html](http://www.stat.fi/til/kuol/meta_en.html) [Accessed 27 Jun 2016].
- Fagt S, Andersen LF, Anderssen SA, et al. *Nordic monitoring of diet, physical activity and overweight*. Validation of indicators: Nordic Council of Ministers, 2012.
- Millwood IY, Walters RG, Mei XW, et al. Conventional and genetic evidence on alcohol and vascular disease aetiology: a prospective study of 500 000 men and women in China. *Lancet* 2019;393:1831–42.
- Rubin DB. *Multiple imputation for nonresponse in surveys*. New York: John Wiley, 1987: 258.
- Therneau T, Atkinson B, Ripley B. *rpart: recursive partitioning and regression trees*, 2017: 1–11.
- Buuren Svan, Groothuis-Oudshoorn K. Mice: Multivariate Imputation by Chained Equations in R. *J Stat Softw* 2011;45:1–67.
- R Core Team. *R: a language and environment for statistical computing*, 2013.
- Carstensen B. Age-period-cohort models for the Lexis diagram. *Stat Med* 2007;26:3018–45.
- Eekhout I, van de Wiel MA, Heymans MW. Methods for significance testing of categorical covariates in logistic regression models after multiple imputation: power and applicability analysis. *BMC Med Res Methodol* 2017;17:129:017–404.
- Meng X-L, Rubin DB. Performing likelihood ratio tests with multiply-imputed data sets. *Biometrika* 1992;79:103–11.
- Grambsch PM, Therneau TM. Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika* 1994;81:515–26.
- Pearl J. *Causality: models, Reasoning, and inference*. Cambridge University Press, 2000.
- Tikka S, Karvanen J. Identifying Causal Effects with the R Package causaleffect. *J Stat Softw* 2017;76:1–30.
- Liaw A, Wiener M. Classification and regression by randomForest. *R news* 2002;2:18–22.
- Ganna A, Ingelsson E. 5 year mortality predictors in 498 103 UK Biobank participants: a prospective population-based study. *The Lancet* 2015;386:533–40.
- Manuel DG, Perez R, Sanmartin C, et al. Measuring burden of unhealthy behaviours using a multivariable predictive approach: life expectancy lost in Canada attributable to smoking, alcohol, physical inactivity, and diet. *PLoS Med* 2016;13:e1002082.
- Foreman KJ, Marquez N, Dolgert A, et al. Forecasting life expectancy, years of life lost, and all-cause and cause-specific mortality for 250 causes of death: reference and alternative scenarios for 2016–40 for 195 countries and territories. *The Lancet* 2018;392:2052–90.
- McNeill A, Brose L, Calder R, et al. *Evidence review of e-cigarettes and heated tobacco products: a report commissioned by public health England*, 2018.
- Goniewicz ML, Gawron M, Smith DM, et al. Exposure to nicotine and selected toxicants in cigarette smokers who switched to electronic cigarettes: a longitudinal within-subjects observational study. *NICOTB* 2017;19:160–7.
- Gotts JE, Jordt S-E, McConnell R, et al. What are the respiratory effects of e-cigarettes?. *BMJ* 2019;1:i5275.
- Nwaneri C, Cooper H, Bowen-Jones D. Mortality in type 2 diabetes mellitus: magnitude of the evidence from a systematic review and meta-analysis. *Br J Diabetes Vasc Dis* 2013;13:192–207.
- Samitz G, Egger M, Zwahlen M. Domains of physical activity and all-cause mortality: systematic review and dose-response meta-analysis of cohort studies. *Int J Epidemiol* 2011;40:1382–400.
- Sun Y-Q, Burgess S, Staley JR, et al. Body mass index and all cause mortality in Hunt and UK Biobank studies: linear and non-linear Mendelian randomisation analyses. *BMJ* 2019;394:i1042.
- Shkolnikov VM, Andreev EM, Jdanov DA, et al. Increasing absolute mortality disparities by education in Finland, Norway and Sweden, 1971–2000. *J Epidemiol Community Health* 2012;66:372–8.
- OECD. *Health at a glance 2017*, 2017: 216.
- McAuley A, Denny C, Taulbut M, et al. Informing investment to reduce inequalities: a modelling approach. *PLoS One* 2016;11:e0159256.