Online Supplement: Additional methods and worked example

Excess risks, expected number of excess deaths, and population attributable fractions

We used a lifetable approach to estimate the impact of exposure interventions. This is a counterfactual approach where we estimate the attributable (excess) cumulative risk of death from lung cancer that is the result from exposure to diesel engine exhaust (DEE), while accounting for competing causes of death using information on (age-specific) cancer-specific and all-cause mortality rates.

We make the simplifying assumption that both exposures and mortality rates are constant over a period of one year from each age, leading to piecewise exponential survival, and define *x(t)* to be a function that returns age-specific exposure levels, which we will call an exposure scenario in what follows. We define S(t,*x(t)*) as the proportion of subjects with exposure scenario *x(t)* that remain at risk of dying from lung cancer at the beginning of the interval starting at age t, with S(0,*x(0)*) = 100% by definition. Under these assumptions and with age- and exposure-scenario specific all-cause and lung-cancer mortality rates equal to m(t,*x(t)*) and l(t,*x(t)*) respectively, the proportion of subjects dying in the 1 year interval from age t is:

ND(t,*x(t)*) = S(t,*x(t)*) * (1-exp(-m(t,*x(t)*)))

and the proportion dying from lung cancer:

NC(t,*x(t)*) = l(t,*x(t)*)) / m(t,*x(t)*)) * ND(t,*x(t)*)

The cumulative proportion of the population that has died from lung cancer up until age T, i.e. the cumulative mortality risk at age T, is then calculated as:

R(T,*x(T)*) = $\sum_{t=1}^{T} \mathrm{NC}(\mathrm{t}, x(t))$

For risk assessment purposes, typically T = 80 years is used. We define the “unexposed” exposure scenario as $X_0(t) = 0$ for $t \in N_0$ and assume that mortality rates obtained from general population registries can be used to estimate age-specific mortality rates for this “unexposed” scenario (which holds for agents with a relatively low exposure prevalence or weak effect), indicating these with $m_0(t,X_0(t))$ and $l_0(t,X_0(t))$ for all-cause and lung-cancer mortality respectively. To estimate mortality rates for an exposed population we assume a relative rate model, where an exposure-response function is available that allows estimation of age- and exposure scenario-specific lung cancer mortality rate ratios, which we indicate by h(t,X(t)). For any specific exposure scenario X(t) we then define:

$l(t,X(t)) = h(t,X(t)) * l_0(t,X_0(t))$ (for lung cancer mortality)

$m(t,X(t) = m_0(t,X_0(t)) + (h(t,X(t)) – 1) * l_0(t,X_0(t))$ (for all-cause mortality)

For the (common) situation where a cox-regression model is used to estimate the log-hazard ratio (β) as a (linear) function of cumulative DEE exposure, the lung cancer mortality rate ratio function could be defined as:

$h(t,X(t)) = \exp(\beta * \sum_{i=1}^{t} \mathrm{X(i)})$

To estimate the excess risk due to DEE exposure at age T, we subtract the cumulative lung cancer risk in an unexposed population from that in an exposed population:

ER(T,X(T)) = R(T,X(T)) - R(T,X_0(T))

To estimate the average excess risk at age T in a population with *p* varying exposure scenarios (i.e. different exposure levels and occupational histories) with optional weights w_1,..,w_p, we calculate the weighted average of scenario-specific excess risks:

$$\overline{\mathrm{ER}_{\mathrm{pop}}}(T) = \sum_{i=1}^{p} \mathrm{w}_i * \mathrm{ER}(\mathrm{T}, \mathrm{X}_\mathrm{i}) \,/ \sum_{i=1}^{p} \mathrm{w}_i$$

The expected number of excess deaths at age T in a population of size N_{pop} is then $\overline{\mathrm{ERpop}}(T) * \mathrm{N_{pop}}$.

Finally, we estimate the population-attributable fraction (PAF) of lung cancer deaths due to DEE at age T as:

$$\overline{\mathrm{PAF}_{\mathrm{pop}}}(\mathrm{T}) = \sum_{\mathrm{i}=1}^{\mathrm{p}} \mathrm{w_i} * \frac{\mathrm{ER}(\mathrm{T}, \mathrm{X_i})}{\mathrm{R}(\mathrm{T}, \mathrm{X_i})} / \sum_{\mathrm{i}=1}^{\mathrm{p}} \mathrm{w_i}$$

Selection of exposure scenarios

We selected relevant exposure scenarios from job histories that were obtained for the control population of the Synergy lung cancer case-control study. Age-specific occupational DEE exposure was estimated using a recently developed quantitative JEM (DEE-JEM)[1] that was linked to study participant job histories using the ISCO-68 occupational coding classification.

We excluded information from subjects that lived outside the EU (i.e. in either Canada or Russia; n=2,862), were born before 1930 (n=6,143), or were not old enough to provide complete job history information (i.e. were younger than 65; n=8,612). Of the remaining 3,188 subjects, 2,004 (63%) were never occupationally exposed to DEE, leaving 1,184 exposure scenarios available for the lifetable analysis. All exposed scenarios were assigned the same weight, with the summed weight for these scenarios equal to the empirically derived prevalence of diesel exposure (i.e. 0.37).

The prevalence of DEE exposure in the selected set of subjects used to derive these scenarios was notably higher than that used in several recently published risk and/or impact assessments for occupational diesel exposure (37% versus 3.3%-8.4%). Although the difference is likely at least partly due to these latter studies focusing more on high exposure occupations in high risk industries, we evaluated excess risks and population attributable fractions also using a much lower summed weight of 0.05 for the exposed scenarios.

Estimation of age-specific background all-cause and lung cancer mortality rates

Mortality rate information was obtained for men and women combined from the Eurostat website for the year 2008 and was available as the number of deaths and size of the population at the start of the year in approximate 5-year age-categories. To estimate age-specific mortality rates, we fitted a penalized poisson regression spline model, assigning the midpoint for each age-category and transforming estimated probabilities of dying into mortality rates assuming these were constant within a single year.

DEE exposure-response relations with lung cancer mortality

We used the previously published exposure-response relation by Vermeulen et al.[2] that assumes a linear relation between log(RelativeRate) and cumulative exposure (using a 5-year lag) as follows:

log(RelativeRate) = 0.000982 * cumulative DEE exposure (in µg/m^3-years)

We accounted for the lag in the life-table calculations by discarding the last 5 years of exposure in calculating cumulative exposures at each age.

Example calculation

We provide a detailed example to allow interpretation of the calculations used to derive the estimated effect of limiting DEE exposure levels to 1 µg/m^3, assuming that the lifetime prevalence of DEE exposure is as observed in the Synergy study (i.e. 37%).

The risk of having died from lung cancer by the age of 80 in an unexposed population is estimated to be approximately 3.53%. Across the exposure scenarios, and without any of the new regulatory standards in place, the same risk ranges from 3.53% to 91.2%, with a mean of 4.45% and a median of 3.86%. The scenario for which the very high risk was estimated entailed an estimated exposure of 150 µg/m^3 DEE for 41 years, resulting in a cumulative exposure of 6,160 µg/m^3-years at age 80 and thus a relative rate of 420 for lung cancer mortality at that age. The average excess risk can be directly estimated from these numbers as (4.45%-3.53%)*37% = 340 per 100,000, which is within rounding error of what we report in the paper (341 per 100,000). With a working population of 229 million workers, the additional number of deaths from lung cancer in the EU can be estimated as 3400 * 229 = 778,600. The population attributable fraction can be estimated by noting that the 340 additional cases per 100,000 subjects would occur among a total number of cases of 4.45%*37% + 3.35*63% = 3,757 cases per 100,000 subjects, the ratio of which is 9.0% (we report 8.8%).

After regulatory standards have been applied, the maximum exposure levels to DEE are reduced to 1 µg/m^3. Risks across the intervened exposure scenarios range from 3.53% to 3.67%, with a mean and median of 3.60%. Following the same logic and calculations as before, the average excess risk now is (3.60%-3.53%)*37% = 26 per 100,000, the additional number of deaths in the EU is 260 * 229 = 59,540, and the PAF is 26/3443 = 0.8%.

1 Ge C, Peters S, Olsson A et al. Diesel Engine Exhaust Exposure, Smoking, and Lung Cancer Subtype Risks: A Pooled Exposure-response Analysis of 14 Case-control Studies. *Am J Respir Crit Care Med* 2020.

2 Vermeulen R, Silverman D, Garshick E, Vlaanderen J, Portengen L, Steenland K. Exposure-response estimates for diesel engine exhaust and lung cancer mortality based on data from three occupational cohorts. *Environ Health Perspect* 2014;122:172-7.