APPENDIX 1. Fitted values and attributable fractions

Calculations of fitted values and attributable fractions are readily implemented in statistical packages such as EPICURE, and commonly reported in authoritative reports (16).

A logistic regression model can be fit to the discrete time data in Table 1a. Below we illustrate estimation of fitted values, background, and excess cases, in a discrete time hazards model fitted using PROC NLMIXED.

```
proc nlmixed data=Table1a ;
 parms b0-b4=0;
 odds=exp(b0 + b1*t + b2*w1 + b3*w2 + b4*x);
 p=odds/(1+odds);
 model Y ~ binary(p);
fv=p;                                                  /*Fitted*/
bk=exp(b0+b1*t+b2*w1+b3*w2)/(1+exp(b0+b1*t+b2*w1+b3*w2));   /*Background*/
ex=fv - bk;                                            /*Excess*/
id bk ex;
predict dths out=fit; run;

proc means data=fit noprint; var bk ex; output out=outmn sum= ;
/*values can be summed for records where x exceeds a limit */

proc print data=outmn; run;
```


Richardson DB, *et al. Occup Environ Med* 2020; 77:520–526. doi: 10.1136/oemed-2019-106303

APPENDIX 2. Counterfactual failure times and events

A logistic regression model can be fit to the discrete time data in Table 1b. The model is weighted so that the estimated parameters are based on records for observed person-periods during which workers had consistently conformed to the policy. We use the resultant estimates of the discrete time hazard to calculate counter factual failure times and events under the policy of interest for each cohort member, i, under policy a. Let $S^a(u|\boldsymbol{W}_i)$ denote the probability of surviving through time u (i.e., one minus the probability of being dead by time u). We define $S^a(-1|\boldsymbol{W}_i) = 1$, and then can proceed to calculate survival through each person-period for subject i as one minus the product of the discrete time hazard in each period, $h^a(t|W)$, and the survival through the preceding period,

$$S^a(u|\boldsymbol{W}_i) = 1 - \sum_{v=0}^{u} h^a(v|\boldsymbol{W}_i) S^a(v-1|\boldsymbol{W}_i) L(v),$$

where $L(v)$ equals 1 in our example, given uniform durations of person-periods in our data structure. We can calculate the probability of being dead by time u if policy a had been implemented as, $Y^a(u|\boldsymbol{W}_i) = \sum_{v=0}^{u} h^a(v|\boldsymbol{W}_i)\, S^a(v-1|\boldsymbol{W}_i) L(v)$.

```
proc logistic data=Table1b;
 model Y(event='1')= t w1 w2;
 weight cf;
 output out = table1c p = h; run;

data table1d;
  set table1c;
  by i u ;
  retain S u D u;
  if first.i then do; S_u=1; D_u=0; end;
    delta_D_u = (h * S_u );  /* discrete time hazard of outcome */
    D_u = D_u + (h * S_u );  /* cumulative hazard of the outcome */
  output table1d;
    S_u = 1 - D_u ;  /* probability of surviving */
run;
```

APPENDIX 3. Calculation of counterfactual failure times and events for cause-specific mortality

Suppose there are 2 categories of cause of death of interest, *Y1* and *Y2*, where category *Y2* denotes death due to all causes other than categories *Y1*. Let $h_I^a(t|\boldsymbol{W})$ denote the discrete time hazard rate of outcome *1* at time *t* under policy *a*, and $h_{II}^a(t|\boldsymbol{W})$ denote the hazard of outcome *2* under policy *a*. Under our proposed approach, each hazard function is estimated by a regression model fit to the empirical data for those who comply with the policy. Under this setting of competing risks, survival is calculated simply by extending the expression to include all categories of cause death,

$S^a(u|\boldsymbol{W}_i)=1-\{\sum_{v=0}^{u} h_I^a(v,\boldsymbol{W}_i)S^a(v-1,\boldsymbol{W}_i)L(v)+\sum_{v=0}^{u} h_{II}^a(v,\boldsymbol{W}_i)S^a(v-1,\boldsymbol{W}_i)L(v)\}$. The approach readily extends from two categories to any number of categories (as long as they are mutually exclusive and exhaustive, such that they encompass all causes of death). Below, we illustrate calculations using the SAS package.

```
%let n=2;

%macro models;
%let cnt=0;
data m0; set table1b;
%do %while (&cnt<&n);
proc logistic data=m&cnt;
%let cnt=%eval(&cnt+1);
model Y&cnt(event='1')=t w1 w2; weight cf; output out = m&cnt p = h&cnt ;
run;
%end;
%mend;
%models;

proc sort data=m&n; by i u; run;

data MR MRunit;
set m&n END=EOF ;
by i u ; du=1;
array expi{*} expected_i1-expected_i&n;
array ec{*} ec1-ec&n; array rate{*} h1-h&n;
array hexpi{*} hexpected_i1-hexpected_i&n;
retain S i expected i1-expected i&n  Ec1-Ec&n P i P;
if _n_=1 then do; do a=1 to &n; ec{a}=0; end; P=0; CP=0; end;
if first.i then do; S_i=1; do a=1 to &n; expi{a}=0; end; P_i=0; end;
do a=1 to &n;
```

```
expi{a}=expi{a}+ ( rate{a} * S_i * du); hexpi{a}= ( rate{a} * S_i * du); end;
P_i =P_i + (S_i * du) ;
hP_i = (S_i * du) ;  * hold counterfactual person-period length;
S_i=1-sum(of expected_i1-expected_i&n);
output MRunit;
if last.i then do; do a=1 to &n; ec{a}=ec{a}+expi{a}; end; P=P+P_i; end;
if eof then output MR; run;
```