



Innovations in applied decision theory for health and safety

David B Richardson ¹, Nathan L DeBono ^{2,3}, Colin Berriault,³ Paul A Demers^{2,3}

► Additional material is published online only. To view please visit the journal online (<http://dx.doi.org/10.1136/oemed-2019-106303>).

¹Epidemiology, University of North Carolina-Chapel Hill, Chapel Hill, North Carolina, USA
²Dalla Lana School of Public Health, University of Toronto, Toronto, Ontario, Canada
³Occupational Cancer Research Centre, Ontario Health, Toronto, Ontario, Canada

Correspondence to

Dr David B Richardson, Epidemiology, University of North Carolina-Chapel Hill, Chapel Hill, North Carolina, USA; david_richardson@unc.edu

Received 9 November 2019
 Revised 24 March 2020
 Accepted 18 April 2020
 Published Online First 12 May 2020

ABSTRACT

Objectives There are established methods for occupational epidemiological cohort analysis, such as proportional hazards regression, that are well suited to aetiological research and yield parameter estimates that allow for succinct communication among academics. However, these methods are not necessarily well suited for evaluation of health impacts of policy choices and communication to decision makers. An informed decision about a policy that impacts health and safety requires a valid estimate of the policy's potential impact.

Methods We propose methods for data summarisation that may facilitate communication with managers, workers and their advocates. We calculate measures of effect in a framework for competing events, graphically display potential impacts on cause-specific mortality under policy alternatives and contrast these results to estimates obtained using standard Poisson regression methods. Methods are illustrated using a cohort mortality study of 28 546 Ontario uranium miners hired between 1950 and 1996 and followed through 2007.

Results A standard regression analysis yields a positive association between cumulative radon progeny exposure and all-cause mortality ($\log(\text{RR per } 100 \text{ WLM})=0.09$; $\text{SE}=0.02$). The proposed method yields an estimate of the expected gain in life expectancy (approximately 6 months per worker) and reduction of 261 lung cancer deaths under a policy that eliminated occupational radon progeny exposure.

Conclusions The proposed method shifts attention from covariate-adjusted risk ratios or rate ratios to estimates of deaths that are avoided or delayed under a potential policy. The approach may help inform decision-making and strengthen the connection of epidemiological approaches to data analysis with developments in decision theory and systems engineering to improve health and safety.

Epidemiologists have relatively sophisticated, well-established methods for the analysis of epidemiological cohort data.^{1 2} These methods estimate parameters that may be well suited to aetiological research and communication among academic investigators, such as covariate-adjusted rate ratios and HRs, often expressed in occupational and environmental epidemiology as exposure-response coefficients. However, estimates of covariate-adjusted trends in rate ratios and HRs are not necessarily well suited to answering the types of health and safety questions confronting decision makers. Moreover, some of the calculations commonly made to communicate background and

Key messages

What is already known about this subject?

- Standard regression analysis yields parameter estimates that allow for succinct communication among academics; however, such methods are not necessarily well suited for evaluation of health impacts of health and safety policy choices.

What are the new findings?

- We propose a method to summarise and graphically display potential impacts of policy alternatives on cause-specific mortality and illustrate it using a cohort mortality study of 28 546 Ontario uranium miners hired between 1950 and 1996 and followed through 2007.
- Under a policy that removed occupational exposure to radon progeny:
 - Fourteen thousand and sixty additional person-years of life would be expected over the study period.
 - Two hundred and sixty-one fewer deaths to lung cancer would be expected, while 72 more deaths due to circulatory disease would be expected.

How might this impact on policy or clinical practice in the foreseeable future?

- The proposed methods of data summarisation and communication may facilitate communication with managers, workers and their advocates.

excess numbers of cases may be misleading for decision makers.³

To improve the utility of such empirical evidence, we consider how health and safety decisions are made in industrial settings, and propose an analytical approach that might be better suited, and yield more informative results, for health and safety decision-making. Decision theory was developed during the post-war period as an applied approach to statistics and engineering that could be used in industrial contexts, recognising that decisions needed to be made in situations where sources of variation and uncertainty may affect success or failure.⁴⁻⁶ While foundations of this approach were laid in the 18th century, methods and theory were improved and refined in the years following World War II and brought together in applied decision theory.⁴ Decision theory frames the decision-making process in terms of a lower order activity



© Author(s) (or their employer(s)) 2020. No commercial re-use. See rights and permissions. Published by BMJ.

To cite: Richardson DB, DeBono NL, Berriault C, et al. *Occup Environ Med* 2020;**77**:520–526.

termed alternative evaluation, which requires the decision maker to identify policy options, evidence or data and goals (ie, what outcomes we prefer). These methods have close connection to epidemiological methods, including use of directed acyclic graphs,⁷ stochastic models and sometimes Markov chain Monte Carlo computations,⁸ suggesting a potential bridge between occupational epidemiology and the tools of decision theory used in industrial management settings.⁶ This suggests opportunities to improve communication about health and safety between health researchers and managers. We illustrate the proposed approach using the example of a study of the impact of a potential exposure reduction policy on cause-specific mortality in a cohort of uranium miners, and discuss implementation of the proposed methods using standard statistical software.

METHODS

The first step in applied decision theory is to clearly state the options or choices under evaluation. For example, suppose that a company owner is interested in the impact of a policy that would lead to a change in workers' exposure to an agent. Suppose that policy *A* is to leave conditions unaffected and as they were in the past; and policy *B* is to invest in personal respiratory protection just for workers in areas of a worksite where dust levels exceeded a specified limit, *m*. Any number of proposed policies could be evaluated, but only two will be discussed for simplicity. Applied decision theory, like occupational epidemiology, draws on what we know; this may involve information drawn from external sources (eg, prior findings) as well as empirical data (eg, from an occupational cohort).

Suppose that we have an occupational cohort of *n* individuals. Let $i=1, 2, \dots, n$, index cohort members, and let *u* be index time on study. Let *X* denote the exposure that will be affected by policy *B*, allowing that exposure may be non-monotonic and time varying so that we let $X(u)$ denote the exposure level at time *u*. Let *W* denote baseline covariates (such as age at hire and sex), and let *Y* denote a binary indicator of the outcome of interest.

Here, we focus on a cohort mortality study where the outcome of interest is all-cause mortality and each individual is under follow-up from study entry until death or censoring due to loss-to-follow-up or administrative end of study. Define study entry as time $u=0$, and denote the time from study entry until administrative end of study follow-up as T_i for person *i*. Let \hat{T}_i denote time to last observation for person *i*, which will end at the time of death, loss to follow-up or administrative end of follow-up, whichever occurs first (ie, $\hat{T}_i \leq T_i$).

A discrete time representation of the person-time at risk in the cohort study divides time on study into a sequence of contiguous time intervals of uniform duration. In the resulting data structure, each cohort member, *i*, contributes one row of data for each time period of observation, letting *u* be the index discrete time intervals, from entry time 0 to \hat{T}_i (table 1a). The binary indicator of the outcome of interest takes a value of '0' for each time period of observation for which $u < \hat{T}_i$; at the last period of observation for each person *i*, a value of '1' is assigned if the outcome of interest was observed, while a value of '0' is assigned to censored observations.⁹ This data structure is used in standard approaches to epidemiological cohort analyses and allows for estimation of the effects of exposure on health outcomes using empirical data.⁹

Statistical methods

We wish to consider the question, "What would be the impact on the occurrence of outcome *Y* of a policy that affected exposure

X?" In the following paragraphs, we contrast two statistical approaches to addressing this question.

Classical regression modelling approach

Let $P(Y_{iu}=1)$ denote the probability that person *i* will experience the outcome event of interest during time interval *u*, conditional on their event-free survival up to the start of time interval *u*. By partitioning the study period into a sequence of narrow discrete time intervals, we are able to estimate the probability of the event, over each of these subperiods. Since $P(Y_{iu}=1)$ is bounded by 0 and 1, it can be modelled using the data structure in table 1a as having a dependence on a set of predictors as,

$$\log\left(\frac{P(Y_{iu}=1)}{1-P(Y_{iu}=1)}\right) = \alpha_u + \beta x_{iu} + \gamma_1 w_{i1} + \gamma_2 w_{i2} \dots \text{Equation 1,}$$

where x_{iu} is the exposure variable of interest, and w_{i1} - w_{i2} are covariates.

Such a model is a discrete time approach to estimation of a model similar to a Cox proportional hazards model for continuous time.^{10 11} The model in Equation 1 includes parameters α_u that describe temporal variation in the hazard function; in practice, temporal variation in the baseline hazard is often modelled more parsimoniously as a polynomial or spline function of the time scale. The output from fitting model Equation 1 is often used to communicate with decision makers. For communication with experts, it is standard to report the estimated coefficient, β , which corresponds to the change in log relative hazard per unit of exposure of interest, or the antilog of β which is the covariate-adjusted discrete time HR.⁹

To derive simpler summary statistics that are often desired in presentations and communication with stakeholders, the estimated coefficients from model Equation 1 sometimes will be used to calculate other quantities to express the impact of *X* on *Y* in the cohort. For example, one summary quantity is the number of cases that can be attributed to exposure. This often is calculated based on the predicted values for the outcome that are derived using the estimated parameters for the fitted model and the covariate pattern associated with each record in the person-period file (online supplementary appendix 1). The summation of fitted values over person-periods is reported as the predicted number of events; and, by using the estimated parameter values, multiplied by the appropriate covariate patterns, one can also estimate the background cases as the predicted values for the outcome *Y* in the absence of exposure (ie, at $X=0$), the excess cases (ie, the fitted values minus background cases), and the attributable fraction (excess cases over fitted values).

Counterfactual approach

An exposure that affects mortality will affect the distribution of person-time in the study cohort; therefore, a policy that affects exposure (eg, reducing *X*) will affect the time of onset of *Y* (3.) For example, a policy that reduced hazardous exposure would lengthen the average time to death among a cohort of workers and increase the amount of person-time accrued in a long-term follow-up of those individuals. Consequently, the classical approach described above to calculation of excess cases and attributable fractions, although performed in the epidemiological literature, is liable to be misleading for decision-makers^{3 12 13} because the calculations are based on the observed records in the person-period file (which reflect the person-time accrued under the observed exposure conditions, rather than what would have been observed under a policy that affected exposure, eg, reduced exposure).

We propose an alternative approach that frames the estimate of the impact of a policy that leads to a change in exposure *X* on

Table 1 Two hypothetical workers in a cohort study with follow-up through 1985. One worker retired in 1982, the other in 1975. Discrete time hazard data structures

(a) Structure including records for person-periods through last observation

ID	Year	Age	Follow-up, <i>u</i>	Covariate, <i>w</i>	Annual exposure	Cumulative exposure	<i>Y</i>
1	1970	31	0	1	0.00	0.00	0
1	1971	32	1	1	0.00	0.00	0
1	1972	33	2	1	0.00	0.00	0
1	1973	34	3	1	0.00	0.00	0
1	1974	35	4	1	0.00	0.00	0
1	1975	36	5	1	0.00	0.00	0
1	1976	37	6	1	0.00	0.00	0
1	1977	38	7	1	1.62	1.62	0
1	1978	39	8	1	1.30	2.92	0
1	1979	40	9	1	1.90	4.82	0
1	1980	41	10	1	1.30	6.12	0
1	1981	42	11	1	1.10	7.22	0
1	1982	43	12	1	0.76	7.98	0
1	1983	44	13	1	0.00	7.98	0
1	1984	45	14	1	0.00	7.98	0
1	1985	46	15	1	0.00	7.98	0
2	1967	37	0	1	3.26	3.26	0
2	1968	38	1	1	3.71	6.97	0
2	1969	39	2	1	1.78	8.75	0
2	1970	40	3	1	4.09	12.84	0
2	1971	41	4	1	2.21	15.05	0
2	1972	42	5	1	1.90	16.95	0
2	1973	43	6	1	0.24	17.19	0
2	1974	44	7	1	0.10	17.29	0
2	1975	45	8	1	0.00	17.29	0
2	1976	46	9	1	0.00	17.29	0
2	1977	47	10	1	0.00	17.29	0
2	1978	48	11	1	0.00	17.29	0
2	1979	49	12	1	0.00	17.29	0
2	1980	50	13	1	0.00	17.29	1

(b) Structure extended to span period through administrative end of follow-up for all cohort members

ID	Year	Age	Follow-up	Baseline covariate	Ann exposure	Cumulative exposure	<i>Y</i>	<i>C</i>	<i>F</i>	<i>H(U)</i>
1	1970	31	0	1	0.00	0.00	0	1	1	0.001859682
1	1971	32	1	1	0.00	0.00	0	1	1	0.001997535
1	1972	33	2	1	0.00	0.00	0	1	1	0.002145584
1	1973	34	3	1	0.00	0.00	0	1	1	0.002304581
1	1974	35	4	1	0.00	0.00	0	1	1	0.002475332
1	1975	36	5	1	0.00	0.00	0	1	1	0.002658700
1	1976	37	6	1	0.00	0.00	0	1	1	0.002855612
1	1977	38	7	1	1.62	1.62	0	1	0	0.003067064
1	1978	39	8	1	1.30	2.92	0	1	0	0.003294121
1	1979	40	9	1	1.90	4.82	0	1	0	0.003537928
1	1980	41	10	1	1.30	6.12	0	1	0	0.003799711
1	1981	42	11	1	1.10	7.22	0	1	0	0.004080785
1	1982	43	12	1	0.76	7.98	0	1	0	0.004382559
1	1983	44	13	1	0.00	7.98	0	1	0	0.004706544
1	1984	45	14	1	0.00	7.98	0	1	0	0.005054358
1	1985	46	15	1	0.00	7.98	0	1	0	0.005427736
2	1967	37	0	1	3.26	3.26	0	1	0	0.000000263
2	1968	38	1	1	3.71	6.97	0	1	0	0.000000282
2	1969	39	2	1	1.78	8.75	0	1	0	0.000000303
2	1970	40	3	1	4.09	12.84	0	1	0	0.000000326

continued

Table 1 continued

(b) Structure extended to span period through administrative end of follow-up for all cohort members

ID	Year	Age	Follow-up	Baseline covariate	Ann exposure	Cumulative exposure	Y	C	F	H(U)
2	1971	41	4	1	2.21	15.05	0	1	0	0.000000350
2	1972	42	5	1	1.90	16.95	0	1	0	0.000000376
2	1973	43	6	1	0.24	17.19	0	1	0	0.000000404
2	1974	44	7	1	0.10	17.29	0	1	0	0.000000434
2	1975	45	8	1	0.00	17.29	0	1	0	0.000000466
2	1976	46	9	1	0.00	17.29	0	1	0	0.000000501
2	1977	47	10	1	0.00	17.29	0	1	0	0.000000538
2	1978	48	11	1	0.00	17.29	0	1	0	0.000000578
2	1979	49	12	1	0.00	17.29	0	1	0	0.000000621
2	1980	50	13	1	0.00	17.29	1	1	0	0.000000667
2	1981	51	14	1	–	–	–	0	0	0.000000717
2	1982	52	15	1	–	–	–	0	0	0.000000770
2	1983	53	16	1	–	–	–	0	0	0.000000827
2	1984	54	17	1	–	–	–	0	0	0.000000888
2	1985	55	18	1	–	–	–	0	0	0.000000954

outcome Y in a counterfactual framework.¹⁴ The following steps are taken to make the necessary calculation.

First, an extended data structure (table 1b) is constructed in which we add to the discrete time data structure one record for each person period from \hat{T}_i until T_i such that the extended data structure includes records for each person period from $u=0$ through the administrative end of follow-up, T_i . In addition, for each record in the data structure, we define two indicator variables: (1) a time-varying binary variable, c , that equals 1 for observed person periods, $u \leq \hat{T}_i$, else 0; and (2) a time-varying binary variable, f , that equals 1 for observed person periods, u , in which person i conformed to the exposure level of a given policy, else 0 (eg, for a policy capping exposure below a specified level m , we would assign $f=1$ at all times $t \leq u$ if $X_i(t) \leq m$, else 0).

Second, we fit a weighted logistic regression model with terms to describe temporal variation in the baseline hazard and effects of covariates, where we define the weight as cf denoting the product of c and f . We use the fitted logistic model to generate predicted values of the estimated probability (hazard) of the outcome for each person-period in the data set. Standard statistical packages allow the investigator to output a person-period data structure that now includes the predicted value of the outcome, which is the estimated hazard of the outcome for each record in the person-period file (online supplementary appendix 2). This predicted value is based on the estimated parameters for the weighted regression model where the weighting leads to estimates based solely on the observed data for those workers who conformed to a given policy.

Finally, we use the predicted value of the hazard of the outcome during time interval u to calculate the probability of surviving through that time interval, for each person i and time interval u . Given the extended data structure, we can calculate the survival probability of the outcome to any time or age. The survival probability is calculated as one minus the product of the hazard of the outcome during that time interval and the probability of surviving up to the start of that time interval (online supplementary appendix 2). We define the probability of surviving up to the first time interval as 1 because our estimates are conditional on survival until entrance into the study. Using this information, we can readily calculate the expected risk of death under a proposed policy (such as policy B) for a worker.

The extended data structure in table 1b allows for the expected risk of death to be calculated to any given attained age or time on study.

This approach readily extends to analyses of cause-specific mortality in which we model two (or more) competing risks for mortality (online supplementary appendix 3).

Graphical and tabular communication

The observed and counterfactual person-time and numbers of cause-specific deaths can be used as the basis for graphs, charts, figures and tables that illustrate comparative survival, absolute numbers of events, years of life lost and risk difference in the study population under alternative policies, standardised to the baseline covariate distribution of the cohort. These values can be summed over all records to yield estimates of the overall number of deaths and years of life in the cohort through the administrative end of follow-up under contrasting policies A and B . Projections regarding future events, such as the expected numbers of events if the cohort were followed to a specified date in the future, can be generated by extending the data structure to include additional records for person-periods that span the desired time period.

Empirical example

We focus on a cohort of male miners working underground in hardrock mines in the province of Ontario, Canada. Miners are exposed to a complex mix of airborne hazards that can lead to acute effects as well as chronic diseases. Annual exposure to radioactive decay products of radon is measured in units of working level months (WLM).

First, we fit a standard regression model and report the distribution of person-time and observed, fitted background, and fitted excess cancer deaths, and attributable fraction of cancer deaths. Next, we consider a policy B that removes exposure (ie, 0 WLM per year) using the time-varying exposure compliance variable f . Using the approach described above, we fit a model with terms for year of birth and attained age, and estimate survival times and deaths under the policy cap.

We use the resultant values as the basis for charts, figures and tables that illustrate comparative survival, absolute numbers of

Table 2 Parameter estimates, fitted values, background, excess and attributable fraction of deaths obtained by fitting a logistic regression model to the discrete time cohort data

Parameter	Parameters		Fitted values
	Estimate	SE	
Intercept	-5.62	0.03	
Age*	0.074	0.00	
Age ²	0.0001	0.00	
Age ³	-0.00004	0.00	
Born ≤1925	0.31	0.03	
1925<born ≤ 1935	0.102	0.03	
Cumulative exposure/100†	0.0934	0.0231	
Background cases (BK)			8130
Excess cases (EX)			216
Attributable fraction (EX/(EX+BK))			0.03

*Attained age up to 85 years.

†Cumulative working level months under a 10-year lag.

events, years of life lost and risk difference in the study population under alternative policies. We further extend the calculations to examine cause-specific mortality and illustrate comparative cause-specific survival, rank ordering of leading causes of death and changes in the distribution of deaths due to each cause under different policies.

RESULTS

A standard regression analysis of the cohort data yields an age and birth cohort conditional estimate of the association between cumulative exposure and all-cause mortality, with the estimated coefficient 0.09 (SE=0.02) (table 2). If we use the model coefficients, we calculate 216 excess deaths among those who were exposed, yielding an estimate of 3% attributable fraction (table 2).

Next, consider policy B that removes exposure (ie, 0 WLM per year). Table 3 reports estimates of the impact of such a policy. Under this policy, 8176 deaths would be expected, as compared with the 8346 deaths observed (table 3) by age 85 years. This suggests 175 excess deaths, in contrast to the 216 excess deaths suggested by the calculation in table 2. The excess deaths divided by the total observed deaths suggests an attributable fraction of deaths of about 2%, smaller than the value of 3% reported in table 2.

However, eliminating radon progeny exposure leads to a greater longevity among the workers. The expected number of person-years observed under policy B is 14 060 person-years greater than the number observed in the cohort under policy A (table 3). Under policy B, deaths tend to occur at older ages and there is an increase in years of life expected overall in the cohort

Table 3 Expected deaths in the absence of exposure, excess deaths and attributable fraction of deaths based on the counterfactual failure times. Cohort of underground miners

Observed (policy A)		Expected under policy B		Difference (B-A)		
Person-years	Deaths	Person-years	Deaths	Person-years	Deaths	
998, 510	8346	1 012 570	8171	14 060	-175	
	Circulatory disease	2711	Circulatory disease	2783	Circulatory disease	72
	Cancer excluding lung	1508	Cancer excluding lung	1555	Cancer excluding lung	47
	Lung cancer	1230	Lung cancer	969	Lung cancer	-261
	COPD	334	COPD	348	COPD	14

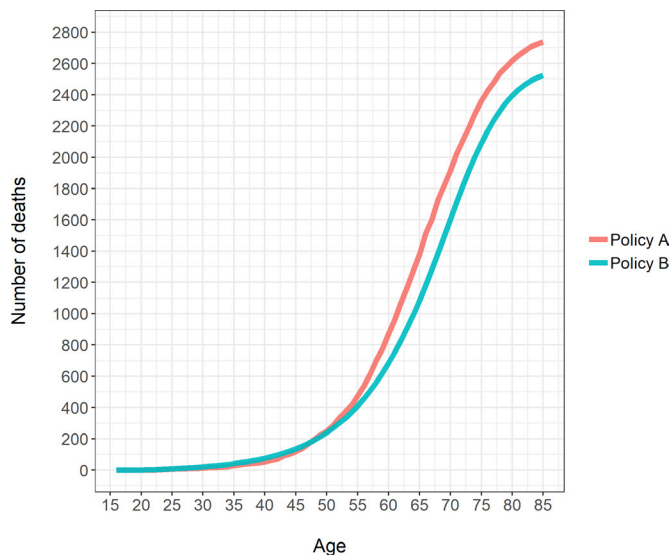


Figure 1 Cumulative mortality by attained age under policy A (red line) and B (blue line).

(approximately six additional months of life per worker in the cohort).

Figure 1, which illustrates a comparative survival curve, shows that by attaining age 85 years, the percentage of workers dead under policy B is lower than under policy A. Online supplementary figure 1 illustrates the gain in years of life by categories of attained age. The bar chart illustrates the gain in the number of person-years of life in each category of attained age expected in the cohort under policy B as compared with policy A. The bars extend towards positive numbers in each category of attained age, indicating that if the occupational health and safety policy B had been in place, there would be more workers alive at each attained age. Most of the years of life gained under policy B occur among people in the 45–54 and 65–74 years of attained age categories.

Table 3 reports cause-specific mortality for lung cancer and other causes under policies A and B. The number of lung cancer deaths is reduced by 261 deaths under policy B compared with policy A; in contrast, the numbers of deaths due to other circulatory disease is increased by 72 deaths. Figure 2 illustrates that policy B shifts the distribution of deaths between categories of cause of death. Online supplementary figure 2 illustrates the cause-specific cumulative mortality functions and shows that under policy B the cumulative incidence of lung cancer (red line) does not increase as steeply as under policy A, but the cumulative incidence of other causes of death may increase in its stead. Table 3 may help a decision maker appreciate the shifting

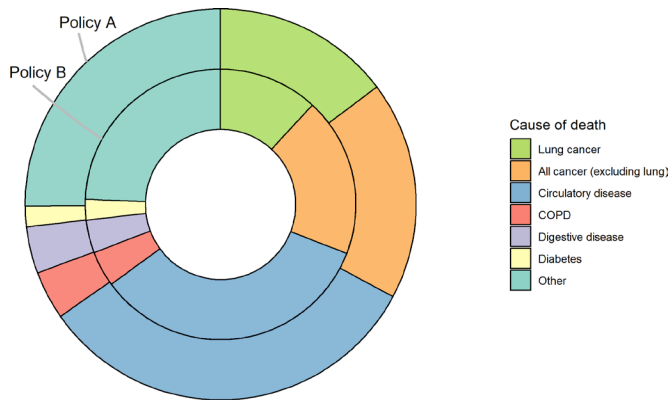


Figure 2 Shifting causes of death. Distribution of deaths under policy A and under policy B by cause of death: policy A (observed); policy B (expected).

distribution of the expected rank order of leading causes of death under different policy options.

DISCUSSION

This paper focuses on methods for occupational cohort mortality analysis that may improve communication with occupational safety and health decision makers. These methods are framed in terms of evaluations of policies on the occurrence of health outcomes; we propose a method to generate failure times and expected events under different occupational policy options. These counterfactual failure times may be compared with observed failure times; a gain in expected years of life may be observed if a policy reduces a hazardous occupational exposure. The approach uses the information generated by standard statistical packages and therefore is readily implemented. We illustrate how these generated failure times and cause-specific expected events can be used for communicating about the potential impact of policies that affect occupational exposures.

As illustrated by our empirical example, the proposed calculations emphasise the limited utility of evaluations of health and safety policies in terms of an expected change in the overall number of deaths. A health and safety policy will not prevent death; however, it might delay when death occurs and shift the distribution of causes of death. This is an important point for communication with decision makers, and frames the benefit of a reduction in harmful exposure in terms of a loss (or gain) of person-time, and a shift in the distribution of causes of death. Of course, applications of the proposed method to different occupational hazards, and different policy interventions, could lead to different patterns of mortality.

The proposed method overcomes a technical limitation of standard approaches to calculation of attributable risks. One cannot consistently estimate the number of deaths expected under policy B simply by multiplying the expected mortality rate under policy B by the observed person moments in the cohort under policy A; and a classical estimate of the excess deaths (ie, the background and excess calculated from the fitted values outputted by a Poisson regression model) may lead to a distorted impression of the impact of a policy under consideration on all-cause mortality.³ The proposed method employs a counterfactual approach to such calculations. Counterfactuals are, by definition, unobserved; yet the approach offers a framework for addressing questions inherent in risk assessments regarding the potential benefit or harm of a (counterfactual) scenario in which an environmental or occupational agent was affected.¹⁵

Of course, the proposed approach requires a number of assumptions to yield valid inference. Correct model specification is necessary under the proposed method, just as it is under standard regression analysis. In order for the predicted hazards generated by the proposed method to serve as appropriate estimates of the expected hazards in the cohort under compliance with a specific policy, we require the assumption of no unmeasured confounding (or ‘conditional exchangeability’). Also, like other regression methods, the proposed method requires that we have observed information on the hazard of the outcome among those who complied with the policy at the covariate patterns observed among those who did not comply. The information available for those who complied with a proposed policy serves as the basis for estimation of the hazard of the outcome that would have been observed in the total cohort under a policy that ‘set’ everyone’s exposure to the conditions of compliance. While these are strong assumptions necessary for valid conclusions from the proposed method, they are general conditions necessary for valid inference that apply to classical regression approaches as well. Finally, we have not addressed complexities that arise if a policy intervention leads to subsequent changes in employment status; however, it is possible to consider extensions of this approach to allow employment dynamics to be influenced by policy interventions.

Standard epidemiological models yield conditional measures of association and statistical tests of the null. However, often what decision makers want are estimates of the impact of a policy that are marginal, rather than covariate conditional, and comparisons of explicit policy alternatives, not comparisons to the null. The proposed methods yields such marginal estimates of the impact of a policy or intervention, and the results pertain to a specified target population. We illustrate calculation of potential years of life gained and shifts in numbers of deaths by a specified attained age (figures 1 and 2); each measure has an appeal in terms of communication and may be helpful for consideration by stakeholders when evaluating policy options.

Decision-making requires the explicit statement of policy options under consideration, as well as values that inform selection of the best outcome, because there is inevitably a trade-off among different consequences of a policy decision. A decision-maker may assign values on quality of life or disability associated with different outcomes; however, in the current paper we focus on estimation of health endpoints without attaching values to them.

Applied decision theory offers a potential bridge for epidemiologists who generate information relevant to health and safety decision-making. The proposed approach to estimation of simple measures of the potential impact of occupational health and safety policies may strengthen communication with decision makers, while at the same time improving the validity of such estimates. We have illustrated how some useful perspectives on the impacts of health and safety interventions can follow from the proposed approaches for summarisation and communication of findings from occupational cohort studies.

Contributors DBR conceived of the project. NLD and CB conducted analyses. PAD is the principal investigator of the cohort study and contributed to interpretation of results and writing text. This paper is not being considered for publication elsewhere.

Funding This study was sponsored by the Alpha Foundation for the Improvement of Mine Safety and Health (ALPHA FOUNDATION). The views, opinions and recommendations expressed herein are solely those of the authors and do not imply any endorsement by the ALPHA FOUNDATION, its directors and staff. The Occupational Cancer Centre is funded by the Canadian Cancer Society, the Ontario Ministry of Labour and Cancer Care Ontario.

Competing interests None declared.

Patient consent for publication Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement No data are available.

ORCID iDs

David B Richardson <http://orcid.org/0000-0001-8550-0212>

Nathan L DeBono <http://orcid.org/0000-0002-5087-107X>

REFERENCES

- 1 Breslow NE, Day NE. *Statistical methods in cancer research: the design and analysis of cohort studies*. International Agency for Research on Cancer: Lyon, 1987.
- 2 Kleinbaum DG, Klein M. *Survival analysis: a self-learning text*. New York, NY: Springer, 2005.
- 3 Richardson DB, Keil AP, Cole SR, *et al*. Observed and expected mortality in cohort studies. *Am J Epidemiol* 2017;185:479–86.
- 4 Chernoff H, Moses LE. *Elementary decision theory*. New York: Wiley, 1959.
- 5 Hooker CA, Leach JJ, McClennen EF. *Foundations and applications of decision theory*. Dordrecht: Boston: D. Reidel Pub. Co, 1978.
- 6 Stevenson WJ. *Operations management*. 9th ed. Boston: McGraw-Hill Irwin, 2007.
- 7 Rothman KJ, Greenland S, Lash TL. *Modern epidemiology*. 3rd edn. Philadelphia: Lippincott Williams & Wilkins, 2007.
- 8 Hamra G, MacLehose R, Richardson D. Markov chain Monte Carlo: an introduction for epidemiologists. *Int J Epidemiol* 2013;42:627–34.
- 9 Richardson DB. Discrete time hazards models for occupational and environmental cohort analyses. *Occup Environ Med* 2010;67:67–71.
- 10 Breslow NE, Lubin JH, Marek P, *et al*. Multiplicative models and cohort analysis. *J Am Stat Assoc* 1983;78:1–12.
- 11 Prentice RL, Gloeckler LA. Regression analysis of grouped survival data with application to breast cancer data. *Biometrics* 1978;34:57–67.
- 12 Greenland S, Robins JM. Conceptual problems in the definition and interpretation of attributable fractions. *Am J Epidemiol* 1988;128:1185–97.
- 13 Greenland S, Drescher K. Maximum likelihood estimation of the attributable fraction from logistic models. *Biometrics* 1993;49:865–72.
- 14 Richardson DB, Cole S, AP K. Standardizing discrete time hazard ratios with a disease risk score. *Am J Epidemiol* 2020.
- 15 Flanders WD, Klein M. Properties of 2 counterfactual effect definitions of a point exposure. *Epidemiology* 2007;18:453–60.