ORIGINAL ARTICLE

Estimation of benefit of prevention of occupational cancer for comparative risk assessment: methods and examples

Lukas Jyunh-Hsiarn Lee,1,2,3,4 Yu-Yin Chang,1 Saou-Hsing Liou,1 Jung-Der Wang4,5

ABSTRACT

Objectives To quantify the life years gained and financial savings by preventing a case of occupational cancer.

Methods The authors retrieved data from the Taiwan Cancer Registry and linked them with the National Mortality Registry to estimate the survival functions for major occupational cancers: lung, pleural mesothelioma, urinary bladder and leukaemia. Assuming a constant excess hazard for each type of cancer, the authors extrapolated lifetime survival functions by the Monte Carlo method. For each patient with cancer, the authors simulated an age- and gender-matched person without cancer based on vital statistics of Taiwan to estimate life expectancy and expected years of life lost (EYLL). By using the reimbursement data from the National Health Insurance Research Database, the authors calculated the average monthly healthcare expenditures, which were summed to estimate the lifetime healthcare expenditures after adjusting for the corresponding monthly survival probability.

Results A total of 51 408, 136, 12 891 and 5285 new cases of lung, pleural mesothelioma, bladder and leukaemia cancers, respectively, were identified during 1997–2005 and followed until the end of 2007. The EYLL was predicted to be 13.7±0.1, 18.9±0.7, 4.7±0.3 and 19.4±0.5 years for these cancers, respectively, and the lifetime healthcare expenditures with a 3% annual discount were predicted to be US$22 359, US$14 900, US$51 987 and US$59 741, respectively.

Conclusions The burden of these occupational cancers, in terms of EYLL and lifetime healthcare expenditures, was substantial. Such estimates may provide useful empirical evidence for comparative risk assessment that can be applied in health policy-making and clinical decision-making.

INTRODUCTION

Cancer is a major public health issue because of the associated reduction in quantity and quality of life in patients and financial burden placed on patients, their families and the society. In the study of the aetiology of human cancer, research in occupational and environmental epidemiology has identified risk factors of cancer that may be of practical implications for prevention. A WHO study found that approximately 19% (12%–29%) of all cancers were estimated to be attributable to the environment, resulting in 1.3 million deaths each year worldwide.1 The International Agency for Research on Cancer has evaluated over 900 potential carcinogens in the past 4 decades, documenting the workplace as a major source of exposure. Occupational cancers are potentially preventable,2 and the burden of occupational cancer has been estimated in terms of attributable fractions, in the range of 2%–8%.3–5 These estimates imply that a substantial percentage of cancer occurrences could be possibly prevented.

Among various definitions of risks,6 the British Standard 18004:2008 describes it as composed of two factors: likelihood of a hazardous event or exposure(s) and the consequence or severity of injury or ill health that can be caused by the event or exposure(s).7 It provides us a basis to quantify the impact of risk by taking account of both likelihood and consequence of the event.8 Nevertheless, the consequences or outcomes generally cannot be directly compared. To have policy implications on the potential impact, a common unit of measurement of outcomes and benefits is needed. Ideally, the valuation of the human health benefits gained from environmental regulations or public health programmes would include all costs to society, including medical costs, work-related costs, educational costs, the cost of support services required by medical conditions and/or the willingness of individuals to pay to avoid the health risks. Direct medical costs, as suggested by the US Environmental Protection Agency, often provide a lower bound estimate of the costs of illness.9 For comparative risk assessment, we proposed the use of life years or quality-adjusted life years8 gained
and financial savings on healthcare expenditures as the metric of outcome evaluation for comparison with other benefits gained through policy investment.

However, there is a general lack of population-based information on the quantification of health benefits associated with the prevention of occupational cancer. In addition, the investigation of overall medical resource utilisation and economic burden of cancer at a national level is still limited.\textsuperscript{10–13} The objective of this study is to quantify the life years gained and the potential financial savings from preventing a case of occupational cancer.

**METHODS**

This study began after the approval of the Institutional Review Board of the National Taiwan University Hospital (NTUH IRB no. 2008-0829R). We used two different datasets to cross-validate the estimation of the survival functions in this study: the Taiwan Cancer Registry (TCR) and the National Health Insurance Research Database (NHIRD).

**Taiwan Cancer Registry**

We retrieved data from the TCR and linked them with the National Mortality Registry to estimate the survival functions for major occupational cancer. The cancer sites of interest in this study included lung (ICD-9-CM code: 162), pleural mesothelioma (ICD-9-CM code: 165), urinary bladder (ICD-9-CM code: 183) and all leukaemia cases other than chronic lymphocytic leukaemia (ICD-9-CM codes: 204.0, 205.0, 205.1). Only patients aged 18 years or older with histopathological evidence of the disease were included in this study. A total of 69,720 patients were identified with the diagnoses of four major occupational cancers from 1 July 1997 to 31 December 2005. The dataset was then linked with the National Mortality Registry database to verify the vital status for each case until the end of 2007.

**National Health Insurance Research Database**

The reimbursement data of the NHIRD from 1997 to 2007, which contained data for all outpatients and inpatients with diagnoses involving the major occupational cancers, were also utilised. Taiwan established a single-payer National Health Insurance (NHI) programme in 1995. At the end of 2008, the NHI covered 99.0% of Taiwan’s population of 23 million. The NHIRD consists of original claims data for reimbursement and a registry of all enrollees, which allowed researchers to trace all medical services received by enrollees under the NHI programme. To secure the protection of personal confidentiality, all the personal identification numbers were encrypted, and the reimbursement data were transformed into a research database, namely the NHIRD, which were regularly maintained by the National Health Research Institutes. Taiwan’s NHI covers almost all healthcare services for catastrophic illnesses, except some novel technologies and medications for which there is no evidenced-based consensus on their effectiveness. As all types of cancer can be registered as a catastrophic illness and treated without copayments, the criteria for the registration were very strict and established by the medical specialty boards and the Bureau of NHI. Therefore, we only included the four types of cancer cases (as described above) that were successfully registered in the dataset of catastrophic illnesses from 1 July 1997 to 31 December 2007. To ensure that all cases were incident cases during the study period, we excluded the potentially prevalent cases that were admitted in any hospital with a diagnosis of any of the four types of cancer before 1 July 1997.

**Extrapolation method to obtain lifetime survival functions**

Monte Carlo simulation was used to extrapolate survival for up to 50 years to derive the lifetime survival function after the diagnosis of each cancer. Briefly, the survival function for an age- and gender-matched reference population was generated using the Monte Carlo method from the life tables of the general population of Taiwan in the corresponding year. The lifetime survival of the patients with cancer (up to 50 years) was obtained using linear extrapolation of a logit-transformed curve of the survival ratio between the cancer cohort and reference population, under the assumption of a constant excess hazard model.\textsuperscript{14,15} The expected years of life lost (EYLL) for a specific disease was defined as the lifetime survival difference between the disease cohort and an age- and gender-matched reference population. We estimated the average EYLL by calculating the difference in the areas under the long-term survival curves between the cohort of cancer patients and the age- and gender-matched reference population. To facilitate the estimation, we used the ISQOL software, which was built in the R statistical package and can be freely downloaded from http://www.stat.sinica.edu.tw/~jshwang.

**Measurement of lifetime costs paid by the NHI**

We used the NHIRD data to establish the cohorts of incident cancer patients for the estimation of the lifetime healthcare expenditures from the perspective of the NHI. The total lifetime cost of a patient refers to all the direct healthcare expenditures paid by the NHI from the date of cancer diagnosis until the date of death. Patients that were alive at the end of 2007 were censored. The dates of death were retrieved from hospitalisation files and the catastrophic illness registration files. By retrieving the reimbursement data from the NHIRD, we were able to calculate the average healthcare expenditures spent by the patients with different types of cancer at time $t$, which can be summed for the lifetime after adjusting for the corresponding survival probability at time $t$. The effective sample size in each month in the follow-up period was applied for the calculation of the average monthly healthcare expenditure using the SAS software. The lifetime healthcare expenditure per case was estimated by multiplying the monthly mean survival probability with the corresponding average monthly healthcare expenditure, adjusting for the annual discount rate.

**Validation of the Monte Carlo extrapolation and comparison with the Kaplan–Meier method**

Empirical data from the TCR and NHIRD provided an opportunity to cross-validate the estimation of survival functions and the actual performance of the Monte Carlo extrapolation. We first included a subcohort of patients diagnosed as having the cancers of interest between 1 July 1997 and 31 December 2001. The subcohorts were only followed until the end of 2001 or for a period of 4.5 years. Then, we extrapolated the data through the end of 2007 or for an additional period of 6 years, using the Monte Carlo method to estimate mean survival months. For the cohorts that were followed to the end of 2007, the Kaplan–Meier method was applied to calculate the mean survival month based on a follow-up of 10.5 years as the ‘gold standard’ for comparison. We presented the relative biases for each cancer to show the differences between the Kaplan–Meier estimates and those of the Monte Carlo extrapolation method for both the TCR and the NHIRD. Because of the limited sample size (N=136) of pleural mesotheliomas, we only performed validation for lung cancer, bladder cancer and leukaemia patients.
Methodology

Table 1 Characteristics of cancer cohorts and estimated lifetime healthcare expenditures based on TCR (July 1997—December 2007) and NHIRD during 10.5 years of follow-up (July 1997 to December 2007)

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Lung cancer</th>
<th>Mesothelioma</th>
<th>Bladder cancer</th>
<th>Leukaemia</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>TCR</td>
<td>NHIRD</td>
<td>TCR</td>
<td>NHIRD</td>
</tr>
<tr>
<td>Number</td>
<td>51,408</td>
<td>68,926</td>
<td>136</td>
<td>428</td>
</tr>
<tr>
<td>Age (years), mean±SD (range)</td>
<td>67.7±12.0</td>
<td>67.6±12.2</td>
<td>60.0±14.9</td>
<td>64.3±14.9</td>
</tr>
<tr>
<td>Median survival (month)</td>
<td>8.1</td>
<td>8.3</td>
<td>6.0</td>
<td>8.2</td>
</tr>
<tr>
<td>Male, %</td>
<td>68.6</td>
<td>66.5</td>
<td>71.3</td>
<td>63.8</td>
</tr>
<tr>
<td>Life expectancy (years)</td>
<td>2.4±0.1</td>
<td>2.4±0.1</td>
<td>2.5±0.7</td>
<td>3.5±0.7</td>
</tr>
<tr>
<td>EYLL</td>
<td>13.7±0.1</td>
<td>13.8±0.1</td>
<td>18.9±0.7</td>
<td>15.2±0.7</td>
</tr>
<tr>
<td>Lifetime healthcare expenditures (US$), mean±SE</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Discount 3%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>After cancer diagnosis</td>
<td>22,359±2033</td>
<td>22,483±2094</td>
<td>14,900±2844</td>
<td>18,856±2720</td>
</tr>
<tr>
<td>Included duration 3 months before diagnosis</td>
<td>22,667±1971</td>
<td>22,791±2033</td>
<td>15,209±2844</td>
<td>19,165±2720</td>
</tr>
<tr>
<td>Discount 5%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>After cancer diagnosis</td>
<td>22,051±2033</td>
<td>22,236±2033</td>
<td>14,776±2535</td>
<td>18,609±2782</td>
</tr>
<tr>
<td>Included duration 3 months before diagnosis</td>
<td>22,359±1971</td>
<td>22,483±1971</td>
<td>15,023±2535</td>
<td>18,918±2720</td>
</tr>
</tbody>
</table>

1US$ = 29.322 New Taiwan Dollar in 2010. EYLL, expected years of life lost; NHIRD, National Health Insurance Research Database; TCR, Taiwan Cancer Registry.

Statistical analysis

The statistical analyses were performed using the SAS software, V9.2 (SAS Institute Inc.). To apply Monte Carlo simulation, we used the ISQoL software (available from http://www.stat.sinica.edu.tw/~jschwang) for the estimation of lifetime survival functions and the calculation of EYLL and lifetime healthcare expenditures. We also performed sensitivity analyses considering an annual discount rate of 3% and 5% and with a disease duration that included the 3 months before the date of cancer diagnosis.

RESULTS

The descriptive characteristics of the four cancer cohorts are summarised in table 1. The mean age for leukaemia patients is the youngest, which generally occurred in the early 50s, compared with the other three types of cancer, which generally occur at an average age of early to late 60s. As the median survival for lung cancer and mesothelioma are generally <1 year and their life expectancies are usually <4 years, these two types of cancer would save 13–19 life years per case plus the estimated lifetime healthcare expenditures. For lung cancer, the estimates of median survival, EYLL and lifetime healthcare expenditures based on the TCR data were comparable with those based on the NHIRD data (table 1). The cohort of pleural mesothelioma is only composed of 156 patients in the TCR, likely due to the difficulty in diagnosing this cancer solely based on pathology. The number recorded in the NHIRD during the same period was 428. All the estimates were slightly different, although they showed a similar trend. For example, the mean (±SE) lifetime healthcare expenditures based on the TCR after cancer diagnosis was US$15 703 (±2411) at an annual discount rate of 5%, while the mean based on the NHIRD was US$19 598 (±2720). The patients with bladder cancer had the highest life expectancy of 11.3 years, and their lifetime healthcare expenditures generally exceeded US$50 000. Prevention of leukaemia would save an EYLL of 19.4 years per case and a lifetime healthcare expenditure of approximately US$60 000–US$70 000, based on this study. The sensitivity analyses with different discount rates and disease costs, either after cancer diagnosis or with a disease duration that included 3 months prior to diagnosis, did not have a large effect except for urinary bladder cancer, which has a life expectancy of 11.3–11.6 years.

The survivals based on the subcohort of cancer patients established in the 4.5-year period were extrapolated to an additional 6 years using the Monte Carlo method and then compared with actual survival months calculated by the Kaplan–Meier method that used the complete 10.5-year follow-up from July 1997 to December 2007. The relative biases for the two methods are summarised in table 2. The relative biases for the Monte Carlo method were generally below 5% except for leukaemia, which ranged between −10.64% and −12.89%. The differences of relative biases between the TCR and NHIRD for the same cohort and time period were <2.5%, indicating a high agreement between the two datasets. The Monte Carlo method tended to underestimate the long-term survival compared with the Kaplan–Meier method, and the largest relative biases were consistently found in the leukaemia patients.

Table 2 Estimates of mean survival months in 10.5 years of follow-up using the Monte Carlo method based on the first 4.5 years of follow-up data for the TCR and the NHIRD were compared with the Kaplan–Meier estimates based on 10.5 years of follow-up (July 1997 to December 2007)

<table>
<thead>
<tr>
<th>Cancer site</th>
<th>10.5-year follow-up</th>
<th>Extrapolation based on the first 4.5-year follow-up</th>
<th>TCR</th>
<th>Kaplan–Meier method Mean</th>
<th>Monte Carlo method Mean</th>
<th>Relative bias (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lung</td>
<td>21.14</td>
<td>20.07</td>
<td>0.4</td>
<td>−5.06</td>
<td>9.50</td>
<td>28.87</td>
</tr>
<tr>
<td>Bladder</td>
<td>77.81</td>
<td>76.36</td>
<td>1.45</td>
<td>−1.86</td>
<td>10.86</td>
<td>20.80</td>
</tr>
<tr>
<td>Leukaemia</td>
<td>38.04</td>
<td>33.14</td>
<td>1.79</td>
<td>−12.89</td>
<td>10.47</td>
<td>28.32</td>
</tr>
</tbody>
</table>

NHIRD, National Health Insurance Research Database; TCR, Taiwan Cancer Registry.
DISCUSSION

Our study illustrated a practical approach to estimate EYLL and lifetime healthcare expenditures for occupational cancers using data from administrative databases. A particular strength of this study is the use of two separate national population-based databases, the TCR and NHIRD, with adequate sample sizes of cancer patients and a long follow-up period (July 1997 to December 2007). This enabled us to cross-validate the estimates. All life expectancy estimates for the cancers from the two databases were relatively close or within a 1-year difference, implying the validity of datasets and the consistency of estimation method. However, there was a difference of EYLL for mesothelioma estimated from the TCR and NHIRD (18.9–15.2=3.7 years). The difference is likely attributed to the different mean ages at diagnosis, that is, 64.3–60.0=4.3 years. The cancer patients included in the TCR dataset were those with histopathological evidence of disease and were approximately 4 years younger than those registered in the NHIRD, which used clinical data only. Our estimate of lifetime cost to the NHI also corroborates with another study using claims data from a single tertiary care medical centre in Taiwan during 1999–2002; the undiscounted average cost up to 10 years for lung cancer was NT$ 448 371 (US$ = NT$ 35 in 2002), which is equivalent to US$167 083 in 2010. Thus, we tentatively concluded that the two datasets, TCR and NHIRD, and the method used in this study are valid. The resulting estimates may be useful in future health policy decisions, especially regarding cancer prevention programmes.

Based on the results that were estimated from histopathologically diagnosed cancer cases abstracted from the TCR dataset, we found that the successful prevention of a case of lung cancer, pleural mesothelioma, bladder cancer and leukaemia would save US$222 859, US$14 900, US$51 987 and US$59 741, respectively. If a prevention programme costs more than it were estimated to save, the residual money can be divided by the number of life years saved to obtain an estimate of cost-effectiveness or cost per life year gained, which can be directly compared with different clinical treatments, rehabilitation, etc. Moreover, the establishment of national occupational standards for carcinogens and strict enforcement of these standards may provide additional benefits: the savings of human capital loss (indirect costs) due to the occurrence of cancer, the health benefits of preventing non-cancer outcomes related to the same exposures and the avoidance of human suffering among cancer patients and their families.16

The lifetime healthcare expenditures of the patients with bladder cancer were much greater than those with mesothelioma or lung cancer because these patients generally had a longer life expectancy than the others. Avritscher et al17 at the University of Texas MD Anderson Cancer Centre estimated that the average cost for bladder cancer was US$65 158, and the predicted lifetime costs for patients averaged US$120 684 for the best-case scenario and US$99 270 for the worst-case scenario. Longer survival is generally associated with a longer period of surveillance, a higher likelihood of treatment of recurrent tumours and a higher cost of treating with possible complications. The mean estimates of lifetime healthcare expenditures have their variance that is expressed in SE, as shown in the table 1. Because our data did not include the details of sources of variance,9 we simply show figure 1 to illustrate the temporal changes of lifetime healthcare expenditures, using lung cancer as an example, which shows that the cost is generally higher at the time of diagnosis and initial treatment, stabilises after 6–12 months and then is slightly elevated near the end of life. With the advanced development of new technologies, medications and therapies in cancer treatment, the survival of cancer patients will be improved and healthcare expenditures will increase. Thus, the current estimates may be lower than the true value of the healthcare expenditure paid by the NHI in the future.

Limitations

This study has the following limitations that should be addressed. First, we adopted the insurer’s perspective, and only direct medical costs were estimated in this study. Based on the estimates of US National Institutes of Health, the 2007 overall annual costs of cancer were US$226.8 billion; all healthcare expenditures were US$103.8 billion, and indirect morbidity costs and mortality costs accounted for more than 50% of the total costs of cancer.18 Due to the lack of empirical data on the costs of lost productivity due to illness or premature death, our results underestimate the cost of illness to the whole society. Because the estimation of cost of illness is generally a lower bound of that of applying the willingness-to-pay method, it implies that our results leave the true figure to some extent.19 Moreover, since the average healthcare expenditure is much lower compared with all the developed countries,20 it requires appropriate adjustments for the results to be transferable to other countries or jurisdictions. However, since we have quantified the EYLL, one may also use these estimates for calculating the cost-effectiveness ratio for a prevention programme that is able to provide the expected number of prevented cancer cases.

Second, because the data from the NHIRD were collected primarily for health insurance administration use, we are concerned that the follow-up of patient survival status may not be comprehensive. Using the TCR data cross-linked with the National Mortality Registry database as the gold standard, we have validated that there is relatively little bias in calculating the survival functions based on the NHIRD data (table 2).

Third, the data on disease severity, including cancer staging, and a patient’s socioeconomic status were not available in both databases; these factors may influence the prognosis of cancer and lifetime cost estimation. Future studies are needed to address the potential effects of these factors, including establishment of the NHI on survival and life expectancy for communities with different socioeconomic statuses.

Fourth, there were no available data regarding the occupational exposure to carcinogens in the databases that were used in...
this study. Cancers caused by occupational agents tend to affect younger individuals, especially if the initial exposure to the carcinogen occurs early in their working life. In addition to the life shortening effect directly from work-related exposure to carcinogens, potential interactions between occupational factors and personal lifestyles may also contribute to additional life shortening effect, such as synergistic interaction between smoking and asbestos exposure in Americans and Chinese workers. Therefore, EYLL and the cost in occupational cancer will be higher than non-occupational-related cancers. The estimates, based on all cases of cancer, are more likely an underestimation of the prevention benefits of occupational cancers.

CONCLUSIONS

In conclusion, we have proposed a feasible approach for measuring EYLL, as well as the lifetime healthcare expenditures for occupational cancers, based on the analyses of population-based administrative databases. The same method could be applied in future studies to stratify cancer patients according to their cancer stages and evaluate the health benefits of early detection of cancer. This study has provided empirical evidence that the burden from these occupational cancers on EYLL and lifetime healthcare expenditures was substantial, which warrants the attention of policy makers and highlights the importance of investing more resources in prevention programmes.

Acknowledgements We thank the Bureau of Health Promotion, Department of Health of Taiwan, for provision of the data from the Taiwan Cancer Registry. We are grateful that the National Health Research Institutes (NHRI) of Taiwan for funding this research (intramural project EO-100-EQ-PP04) and the Institute of Occupational Safety and Health (IOSH-100-M303) for the support in recognition of occupational cancer and case investigation. The interpretation and conclusions contained herein do not represent those of Bureau of National Health Insurance, Department of Health or National Health Research Institutes.

Contributors LJ-HL, S-HL and J-DW contributed to the conception and design of the study; Y-YC and LJ-HL performed the data analysis. All authors have contributed to interpretation of the results and take public responsibility for its accuracy. Each author is confident in the validity of this work, has reviewed the final version of the manuscript and approves it for submission.

Funding This study was supported by the National Health Research Institutes (NHRI) intramural project EO-100-EQ-PP04, and the Institute of Occupational Health and Safety (IOSH-100-M303). The funder helps in acquisition of data from the National Health Insurance Research Database. The authors have worked on the research independently. The interpretation and conclusions contained herein do not represent those of the National Health Research Institutes.

Competing interests None declared.

Ethics approval Ethics approval was provided by the Institutional Review Board of the National Taiwan University Hospital.

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES