Full Paper

Occupational cancer in Britain
Statistical methodology

Sally J Hutchings*,[1] and Lesley Rushton[1]

[1]Department of Epidemiology and Biostatistics, School of Public Health and MRC-HPA Centre for Environment and Health, Imperial College London, St Mary's Campus, Norfolk Place, London W2 1PG, UK

An approach using the attributable fraction (AF) has been developed to estimate the current burden of occupational cancer in Britain. The AF combines the relative risk (RR) associated with exposure with the proportion exposed. For each cancer–exposure pairing, the RR is selected from key epidemiological literature such as an industry, or population-based study, meta-analysis or review. The CARcinogen EXposure (CAREX) database provides point estimates for the number of workers exposed to a range of carcinogens; alternative sources are national surveys such as the Labour Force Survey and Census of Employment. The number of workers exposed are split between high and low exposure levels matched to appropriate RRs from the literature. The relevant period for cancer development during which exposure occurred is defined as the risk exposure period (REP). Estimation of the numbers ever exposed over the REP takes into account the changes in the number of people employed in primary and manufacturing industry and service sectors in Britain where appropriate, and adjustment is made for staff turnover over the period and for life expectancy. National estimates of the population ever of working age during the REP are used for the proportion denominator. Strategies have been developed to combine exposure AFs correctly while avoiding double counting and minimising bias. The AFs are applied to national cancer deaths and registrations to obtain occupation-attributable cancer numbers. The methods are adaptable for other diseases and other geographies, and are also adaptable to more sophisticated modelling if better exposure and dose–response data are available.

Doll and Peto (1981) estimated the proportion of cancer deaths in Britain due to occupational causes as 4% (with an uncertainty range from 2 to 8%), which equates to ∼6000 deaths per annum (with a range of 3000–12 000). The aim of this project was to update the estimate of the burden of cancer in Britain due to occupation. Several different approaches are available to calculate attributable cancer burden. However, for this study, the occupation-attributable fraction (AF), that is, the proportion of total cancers that would not have occurred in the absence of occupational exposures, was used. This required estimation of the risk of the disease associated with the carcinogen of interest (e.g., the relative risk (RR)) and also the proportion of the population exposed to the carcinogen of interest. In this paper, the methodology developed for this project is described. The methods are illustrated in the accompanying papers in this supplement. The strengths and weaknesses (i.e., sources of bias and uncertainty) of this methodology are discussed, as well as interpretation of the results. More detail of the methodology is provided in the Health and Safety Executive technical report (HSE, 2011).

METHODOLOGY

The overall AF was calculated on a 'cancer-by-cancer' basis. Relative risks were obtained from currently available epidemiological studies that were considered the most appropriate.

The choice of these 'best studies' then determined the method by which AF was calculated.

Identifying the occupational exposures

For each cancer, major literature sources were reviewed to identify relevant occupational exposures, confounders and other non-occupational risk factors, and to obtain information on latency and trends in mortality, incidence and survival. At two international workshops held as part of the project to discuss the methodology (HSE, 2005, 2007), the participants advised that priority should be given initially to International Agency for Research on Cancer (IARC) Group 1 and 2A occupation-related carcinogens and occupational circumstances (e.g., occupation as a painter). Agents or occupations in these IARC groups were included that had either 'strong' or 'suggestive' evidence of carcinogenicity in humans for the specific cancer site, as defined by Siemiatycki *et al* (2004) and subsequent IARC publications. In addition, there had to be substantial existing exposures in Britain and/or cases of cancer still occurring due to past exposures.

It is quite common for workers to be exposed over a working lifetime to several occupational carcinogens able to cause cancer at a single site. It is incorrect to estimate separately the contribution towards the overall AF of these overlapping exposures. These exposures should ideally be treated as a set, and a study sought that has RR estimates for the set as a whole or a study that takes account of the confounding effects and interactions within the set.

*Correspondence: SJ Hutchings; E-mail: s.hutchings@imperial.ac.uk
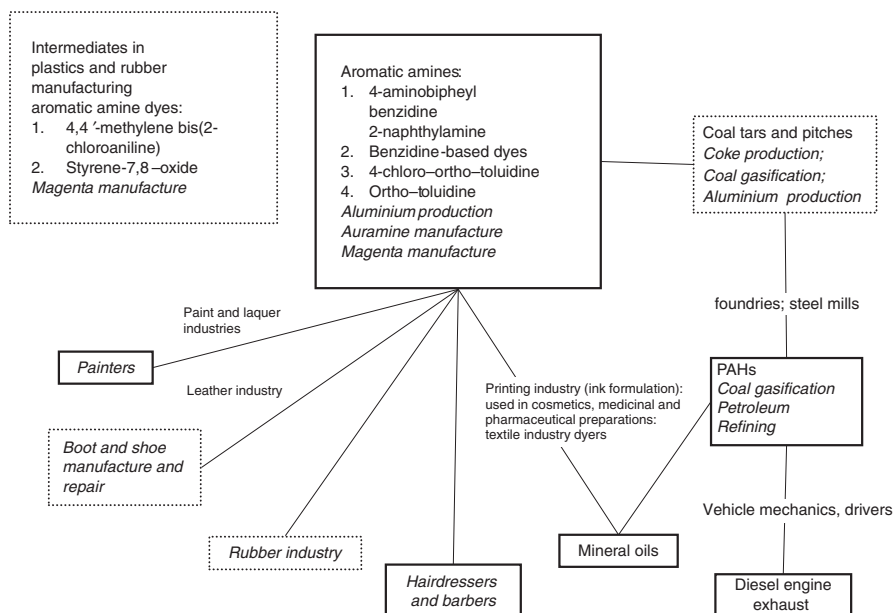
Figure 1 Example of an exposure map for bladder cancer. Carcinogenic agents are in plain type, and exposure scenarios in italics. Lines joining boxes indicate where overlap occurs; industries and occupational circumstances where overlap occurs are indicated in smaller print. For substances and occupations shown in boxes with dotted lines, a separate AF was not estimated, as these exposure scenarios were included with another dominant exposure.

In practice, however, such studies are rarely available. To take account of potential multiple exposures, an 'exposure map' was drawn up for each cancer site (e.g., Figure 1 for bladder cancer) in order to illustrate the potential for double counting of the exposed population, and our strategies to deal with the problem included partitioning exposed numbers between overlapping exposures or estimating only for the dominant carcinogen with the highest risk (Van Tongeren et al, 2012).

For each cancer–exposure pairing, the approach in general was to partition the exposed population between 'higher exposure' and 'lower' and/or 'background exposure' groups, and to apply an appropriate RR from the source study or studies to each group separately.

Choosing RRs

Literature searches were carried out for relevant papers for the selection of RRs. Standard search criteria were used, supplemented by references from review papers and textbook bibliographies. A set of criteria was developed and applied for selecting appropriate RRs from the literature, taking into account the quality of the published study and portability to the British situation. Important issues concerning portability (Steenland and Armstrong, 2006) included: (1) similarity of the published study populations with respect to the nature and levels of exposures – it would be preferable to use British studies or those from countries with a similar industrial heritage and level of development; (2) relevance of the source study exposure period to the target risk exposure period (REP) in the burden estimation (see below); (3) similarity of the distribution of known confounders, for example, smoking, in the published study source population and the British target population.

Time limitations restricted the use of systematic review methods for every cancer site/occupational carcinogen or circumstance. Combined studies (pooled studies and meta-analyses) were thus usually the first choice if available, with UK studies added if necessary. The RRs for use in the calculation of AF were identified and extracted from the chosen studies. Relative risks adjusted for important confounding factors, and for interaction effects with the exposures of interest, were preferred. Where only a narrative review was available, giving a range of risk estimates from several relevant studies, we calculated a combined estimate of the RRs, (i) for homogeneous RRs, based on a fixed-effects model using a precision (inverse variance) weighted average (Breslow and Day, 1989; Rothman and Greenland, 1998), or, (ii) for heterogeneous RRs, based on a random-effects model, using methods described by DerSimonian and Laird (1986). Formal systematic reviews and meta-analyses were carried out to estimate RRs for laryngeal and stomach cancers related to asbestos exposure.

Dose–response risk estimates were generally not available in the epidemiological literature, nor were proportions of those exposed at different levels of exposure over time available for the working population in Britain. However, where possible, risk estimates were obtained for an overall 'lower' level and an overall 'higher' level of exposure to the agents of concern. Where no risk estimate could be identified for very low/background levels of exposure, we estimated a RR for the 'lower-exposed' group by (i) taking the harmonic mean of all the available ratios of 'higher' to 'lower' RR estimates for cancer exposure pairs for which data were available and (ii) applying this average ratio to the 'higher'-level estimates to obtain 'lower'-level RR estimates. If the resulting RR estimate was <1, RR was set to 1.

The REP

We defined the REP, based on cancer latency, as the window of time during which exposure to an occupational carcinogen could result in a cancer being diagnosed or appearing in national mortality or cancer registration records. As there was very little data available on cancer latency for occupational carcinogens, two REPs were assumed for this study: for solid tumours a latency between 10 and 50 years and for haematopoietic malignancies, such as leukaemia, a latency between 0 and 20 years. For deaths in our target year of 2005, this gives a REP of 1956–1995 for solid tumours and of 1986–2005 for haematopoietic malignancies.

Although strictly speaking, the REP for cancer registrations recorded in 2004, the year for which estimation has been carried out, would be 1955–1994, for simplification the years for deaths 1956–1995 have also been used, as the proportion exposed will not be affected. Data specific to these REPs are given in Table 1.

Estimating the proportion exposed

If the RR was extracted from a population-based published study, an estimate of the proportion of the population exposed was taken directly from the study. It would be particularly important in this case that the study population was comparable and portable to Britain. In practice, population-based studies were rarely available.

If the RR was extracted from an industry-based study, national data sources such as the CARcinogen EXposure database (CAREX; Pannett et al, 1998) or the UK Labour Force Survey (LFS, 2009) and Census of Employment (ONS, 2009) were used as a basis for estimating the exposed population. For 139 carcinogens, CAREX gives both the total numbers and the number of people exposed by industrial sector for the period 1990–1993. In practice, for exposures not covered by CAREX, there was rarely a good alternative estimate of the proportions in the workplace exposed; estimates of the number of people employed in a relevant occupation using the narrowest possible definition of this occupation (from the Labour Force Survey) or industry (Census of Employment or
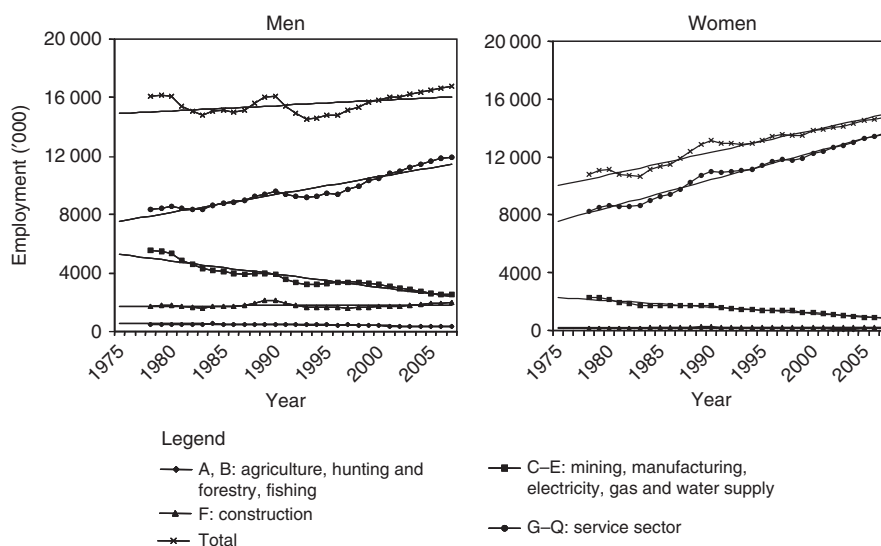
Labour Force Survey) were then used. Exposed numbers from CAREX or national employment sources were allocated to a 'higher'- or 'lower'-exposure group, assuming distributions of exposure and risk that corresponded broadly to those of the studies from which the risk estimates were selected. The initial allocations were based on the judgment of an experienced human exposure scientist; each assessment was then independently peer-reviewed and, if necessary, a consensus assessment agreed (Van Tongeren et al, 2012).

Data from CAREX are not differentiated by sex, and thus they were split between men and women according to an estimate of relative proportions in major Standard Occupational Classification (SOC) groups based on numbers from the 1991 Census. The most often used were: blue-collar workers (SOC major groups 5, 8 and 9) covering skilled trades, shop floor and transport operatives; white-collar workers (SOC major groups 1, 2 and 4) covering managerial, professional, administrative and secretarial; and associate professional, technical, personal and customer service occupations (SOC major groups 3, 6 and 7). In practice, the percentage most appropriate to the particular exposure scenario has been used, with flexibility to use percentages for single rather than grouped SOC major groups if exposed numbers were clearly concentrated in these occupations. Transport operatives (SOC group 8) in the service industries (industrial sectors G-Q) were an example of the appropriate use of a single SOC group for exposure to diesel engine exhaust.

Table 1 Values of variables used in the calculation of attributable fraction specific to the risk exposure periods

Risk exposure period Variable	Solid tumours		Haematopoietic malignancies	
	Men	Women	Men	Women
Latency	10–50 years	10–50 years	0–20 years	0–20 years
Risk exposure period	1956–1995	1956–1995	1986–2005	1986–2005
Peak latency[a]	35 years	35 years	15 years	15 years
Number of years in risk exposure period	40	40	20	20
Total number of those ever of working age during the risk exposure period, and alive in 2005	19.4 million	21.0 million	23.0 million	23.1 million
Ages included	25–90+	25–90+	15–84	15–79

[a]Peak latency is the latency thought to relate to the highest number of cancer cases leading to current deaths or registrations.

Legend
A, B: agriculture, hunting and forestry, fishing
F: construction
Total
C–E: mining, manufacturing, electricity, gas and water supply
G–Q: service sector

Figure 2 Labour Force Survey employment by grouped main industrial sector, A–Q.

Table 2 Factors used for adjusting CAREX exposed numbers to the risk exposure period (solid tumour cancers), and staff turnover factors used in the calculation of numbers ever exposed

Main industry sector		CAREX adjustment factor	Annual Staff Turnover (TO)	Turnover factors based on a 40-year (long latency) and 20-year (short latency) risk exposure period and GB life expectancy tables	
				Long latency risk exposure period	Short latency risk exposure period
Men					
A, B	Agriculture, hunting and forestry; fishing	1	0.07	3	2
C, D, E	Mining and quarrying, electricity, gas and water; manufacturing industry	1.4	0.09	3	3
F	Construction	1	0.12	4	3
G-Q	Service industries	0.9	0.11	4	3
Total		1	0.10	4	3
Women					
A, B	Agriculture, hunting and forestry; fishing	0.75	0.10	4	3
C, D, E	Mining and quarrying, electricity, gas and water; manufacturing industry	1.5	0.14	5	4
F	Construction	0.67	0.15	6	4
G-Q	Service industries	0.8	0.15	6	4
Total		0.9	0.14	6	4

Abbreviations: CAREX = CARcinogen Exposure; GB = Great Britain; TO = turnover.

For an independent estimate of the proportion of the population exposed, the number of workers ever exposed during this period was estimated by extrapolating from a point estimate of exposed workers taken from the period. CAREX data for Britain relate only to the period 1990–1993, and this was the point-estimate year used for these data. For solid tumour cancers where CAREX data were used, adjustment factors were applied to take into account the change in number of people employed in the primary and manufacturing industry and service sectors in Britain between the late 1970s and early 1990s. The factors were estimated from Labour Force Survey employment data, and were grouped by main industrial sector and for men and women separately. The trends on which the factors were based are illustrated in Figure 2, and the factors are given in Table 2. For the Labour Force Survey and Census of Employment, an available year was chosen to represent the number of people employed about 35 years before the target year of 2005, as this was thought to represent a peak latency for the solid tumours, and is also close to the midpoint of the REP for estimating the number of people ever exposed across the period (for which a linear change in employment levels was implicitly assumed). In instances where the Census of Employment was used, the point-estimate data were for 1981 for solid tumours. In instances where the Labour Force Survey was used, the first year available was 1979 for solid tumours, and the data for 1991 was used for short latency malignancies.

Staff turnover factors were applied to the point estimates of the number of people exposed that were specific to the same grouped main industrial sectors as the CAREX adjustment factors. The turnover factors were estimated using Labour Force Survey data for the length of time with the current employer; this was available by length of employment category, and was used to estimate numbers of new starters per year. Ideally, an estimate of staff turnover requires full national starter and leaver data across the REP for all industrial sectors. In the absence of this quality of data, estimating turnover directly using new starters in years within the REPs gives the best approximation for the purpose of estimating those ever exposed. This method estimates starters in the past year as a proportion of the average number of people employed (Gregg and Wadsworth, 2002). Exposure in occupational epidemiological studies is very often defined as that for at least 1 year; we have adapted this method therefore to exclude short-term labour turnover by taking new starters in the past year who are expected to remain employed for at least 1 year as a proportion of all those expected to be employed for at least 1 year. This is estimated as the number of people recorded as employed between 1 and 2 years divided by the total number of people employed for at least 1 year using Labour Force Survey data averaged over the REP.

The number of people 'ever exposed' during the REP, $N_{e(REP)}$, taking into account turnover is as follows:

$$N_{e(REP)} = n_0 + \{n_0 * TO * t\} \tag{1}$$

where n_0 is the number of people employed at the midpoint of the REP, TO is staff turnover per year and t is the number of years in the REP.

Life table estimates of the proportions of the general population surviving to the target year (2005) were used to adjust $N_{e(REP)}$, so that only the ever exposed cohort members surviving to the target year would be counted. This assumed that there was an even distribution of ages across the exposed cohort in its first

year (e.g., 1956 for the solid tumour REP), and that recruitment to the cohort was in the age range of 15–24 only. The equation becomes:

$$N_{e(REP)} = \sum_{i=a}^{i=b} \{l_{(adj15)i} * n_0/(R - 15)\}$$
$$+ \sum_{k=0}^{k=(age(u)-age(1))} \sum_{j=c+k}^{j=d+k} \{l_{(adj15)j} * n_0 * TO/(age(u) - age(l) + 1)\}$$

(2)

where n_0 is the point estimate of the number of people employed at a midpoint of the REP (adjusted for the CAREX data using factors in Table 2).

TO is staff turnover per year (see Table 2),

R is retirement age (taken here as 65 for men, 60 for women),

$l_{(adj15)i}$ is the proportion of survivors up to age i (the l_i) from the sex- and time-appropriate British Life Table data (1980–1982 for the solid tumour REP, 1994–1996 for the short latency REP) (GAD, Interim Life Tables, 2007), adjusted to include only those still alive at the age of 15 years by taking $l_{(adj15)i} = l_i/l_{15}$,

a to b is the age range achieved by the original cohort members by the target year (e.g., $a = 65$, $b = 100$ for the solid tumour standard REP; $a = 35$, $b = 84$ (men, $b = 79$ for women) for the short latency standard REP),

c to d is the age range achieved by the turnover-recruited cohort members by the target year ($c = 25$, $d = 64$ for the solid tumour standard REP; $c = 15$, $d = 34$ for the short latency REP),

$age(u)$ and $age(l)$ are the upper and lower recruitment age limits (15–24 used here).

The last column of Table 2 gives the turnover factors for each major industrial sector obtained by dividing $N_{e(REP)}$ by n_0. These are of similar magnitude to those used in the Global Burden of Disease methodology (Nelson et al, 2005).

For some exposures, an industrial process was known to have changed at a certain time during the REP, or controls had been put in place to reduce exposure. In this case, different RRs were used for each period. To partition the estimate of number of people ever exposed between the two periods, the age ranges c to d in equation (2) were adjusted in the turnover-recruited part of the equation to suit the restricted time period, and members of the original cohort are only counted once, that is, in the first period.

To estimate the proportion of the population exposed, the estimated number of people 'ever exposed' during the REP was then divided by the total population ever of working age during the same period (Table 1). This was based on national population estimates by age cohort in the target year, obtained by summing the age cohorts 25–34 to 90+ for the solid tumour REP and 15–24 to 80–84 for the short latency REP for men (75–79 for women).

Calculating the AF

If the RR was extracted from a population-based study, the estimate of the proportion of cases exposed was extracted directly from the study, and Miettinen's equation (Miettinen, 1974) was used for calculating the AF:

$$AF = Pr(E|D) * (RR - 1)/RR \quad (3)$$

where $Pr(E|D)$ is the proportion of cases exposed (E = exposed, D = case).

For RRs from industry-based studies, with an estimate of the proportion of the population exposed from an independent data source, Levin's equation was used to calculate the AF (Levin, 1953):

$$AF = Pr(E) * (RR - 1)/\{1 + Pr(E) * (RR - 1)\} \quad (4)$$

where $Pr(E)$ is the proportion of the population exposed.

Levin's and Miettinen's equations are equivalent, and can be derived from one another using the definition of RR and Bayes theorem (Benichou, 2001).

Separate AFs were calculated by industry/occupation and for the 'higher' and 'lower'-exposed groups, and then summed to give an overall AF for the specific cancer–exposure pair. The form of the equation used for multiple exposure levels h, separate industries i and time periods j for Levin's estimator was

$$AF = \{\Sigma_h\Sigma_i\Sigma_j Pr(E_{hij})(RR_{hj} - 1)\}/\{1 + \{\Sigma_h\Sigma_i\Sigma_j Pr(E_{hij})(RR_{hj} - 1)\}$$

(5)

Note that in practice separate AFs are estimated by exposure level, industry and time period, but a common denominator must be used that incorporates all the estimates of proportions of the population exposed and RRs, which contribute to the final exposure-AF in each estimate.

For a few cancer–exposure pairings, alternative approaches were used to estimate attributable numbers and hence AF. The AF for mesothelioma was derived directly from several UK mesothelioma studies that suggest that between 96 and 98% of male mesothelioma cases are due to occupational or paraoccupational exposure (e.g., exposure from living near an asbestos factory or handling clothes contaminated because of occupational exposure) (Howel et al, 1997; Yates et al, 1997; Rake et al, 2009). Combining the results from Rake et al (2009) with those from two studies in which results were reported separately for women (Spirtas et al, 1994; Goldberg et al, 2006) gave estimates of 75–90% for women. To allocate total mesotheliomas between separate industries, industry-specific AFs calculated in the usual way were used to estimate the proportions between industries, but mesotheliomas considered to be paraoccupational (6–13% of the total for men, 45–70% for women) were excluded from the allocation. The ratio of asbestos-related lung cancer to mesothelioma deaths has been suggested to be between two-thirds and one (Darnton et al, 2006). We therefore used a ratio of 1:1 mesothelioma to lung cancer deaths for the estimation of the number of lung cancers attributable to asbestos. This takes into account the impact that past levels of asbestos exposure are having on current incidence by the direct link to mesothelioma deaths that are still increasing, whereas lung cancer in general is declining owing to the reduction in smoking. This assumes, however, that lung cancer has a similar pattern of latency to mesothelioma.

Specialised methods were also used to calculate the AF for lung cancer due to radon exposure, based on estimates of the number of people developing lung cancer from exposure to radon in domestic buildings (HSE, 2005).

The risk estimates for occupational exposure to ionising radiation were derived using generalised linear dose–response models of excess RR per unit of cumulative radiation dose from the United Nations Scientific Committee on the Effects of Atomic Radiation (UNSCEAR 2006) (UNSCEAR, 2008), and were therefore applied to cumulative lifetime dose estimated using data from the Central Index of Dose Information (CIDI; HSE, 1998). For aircrew, who are not covered by CIDI, the mean total lifetime radiation dose per pilot was obtained from a large cohort study of European airline pilots (Langner et al, 2004) and combined with the number of people employed, obtained from the British Airways Stewards and Stewardesses Union (BASSA, 2008).

Combining the AFs

A number of strategies were adopted for combining AFs for several occupational exposures causally related to a specific cancer. Attributable fractions cannot generally be summed directly because (1) summing AFs for overlapping exposures (i.e., agents to which the same 'ever-exposed' workers may have been exposed) may give an overall AF exceeding 100% and (2) summing disjoint

(not concurrently occurring) exposures also introduces upward bias. This bias can only be eliminated by using an alternative formulation of Levin's equation, which requires information from the other disjoint exposures in the denominator.

If it could be assumed that overlapping exposures were independent and their joint effect on initiating or promoting cancer was multiplicative, that is, $RR_{(exp\ 1\ and\ exp\ 2)} = RR_{(exp\ 1)} * RR_{(exp\ 2)}$, the AFs for each exposure in the overlapping set were combined into an overall AF for the set by taking the complement of the product of complements (Steenland and Armstrong, 2006):

$$AF_{overall} = 1 - \Pi_k(1 - AF_k) \text{ for the } k \text{ exposures in the set.} \quad (6)$$

This equation is also used for combining disjoint exposure AFs, as it can be shown that this introduces less bias into the results compared with, for example, direct summing of the AFs.

As explained earlier, if it was not appropriate to assume independence of overlapping exposures, a dominant exposure industrial sector was chosen (Figure 1). Equation (6) was then used to combine the now disjoint exposure AFs.

Estimating attributable numbers

Attributable number of deaths and cancer registrations for the most recent year available at the time of the study for a specific cancer site were estimated by multiplying the AF by the total British (England + Wales + Scotland) cancer-specific deaths (2005; ONS, 2006) and registrations (2004; ONS, 2007), respectively. Data were not always available for some short latency malignancy sub-types (e.g., some leukaemia sub-types in the Welsh and Scottish data). In this case, sub-type proportions in English data or other sources were used. For solid tumours, only cancers in people aged 25 years and above were counted, as these could have been initiated during the REPs being considered. For haematopoietic malignancies, only people aged 15 to 20 years beyond retirement age were included (84 for men as the state retirement age was 65 and 79 for women as the state retirement was 60).
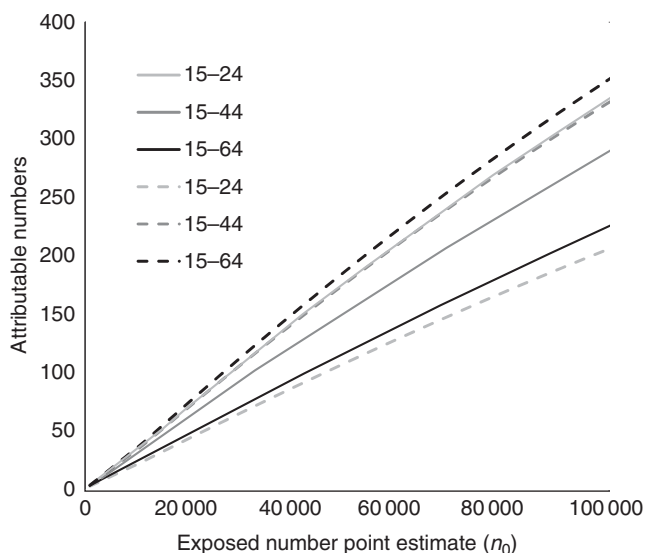
Figure 3 Attributable cancer estimates using a non-age-adjusted attributable fraction (solid lines) and age-adjusted attributable fraction* (dashed lines) and range of turnover cohort recruitment ages, RR = 2. *Results for the 'age-adjusted attributable fraction' are obtained by partitioning $N_{e(REP)}$ from equation (2) by age-group, giving proportions of the population exposed and attributable fraction by age group; these are applied to age-specific cancer numbers to obtain occupation-attributable numbers.

Combining the AFs across cancer sites

To obtain the overall occupation-AF, that is, burden, attributable numbers for each cancer were summed and divided by total cancer numbers. This assumes that there is no overlap between cancer sites and that a single exposure is very unlikely to produce more than one primary cancer in an individual at a specific time. An overall AF for each exposure or each exposure level over all the cancer sites affected by that exposure was estimated in a similar way. It should be noted that attributable numbers should not be summed across different exposures for a specific cancer site; they are estimated by applying the combined AF from equation (6) to totals for that cancer site.

Confidence interval for the overall AF

A 95% random error only confidence interval was calculated for each AF, based only on the variability of the RR estimates used, and using Monte Carlo simulation methods where more than one RR estimate was involved (Greenland, 2004). A total of 100 000 replicates were generated from a log-normal distribution with mean and standard deviation based on the point estimate and standard error for each RR, and then applied to the equation used to calculate the AFs. The 97.5 and 2.5 percentiles from this sample of AFs gave the 95% confidence intervals.
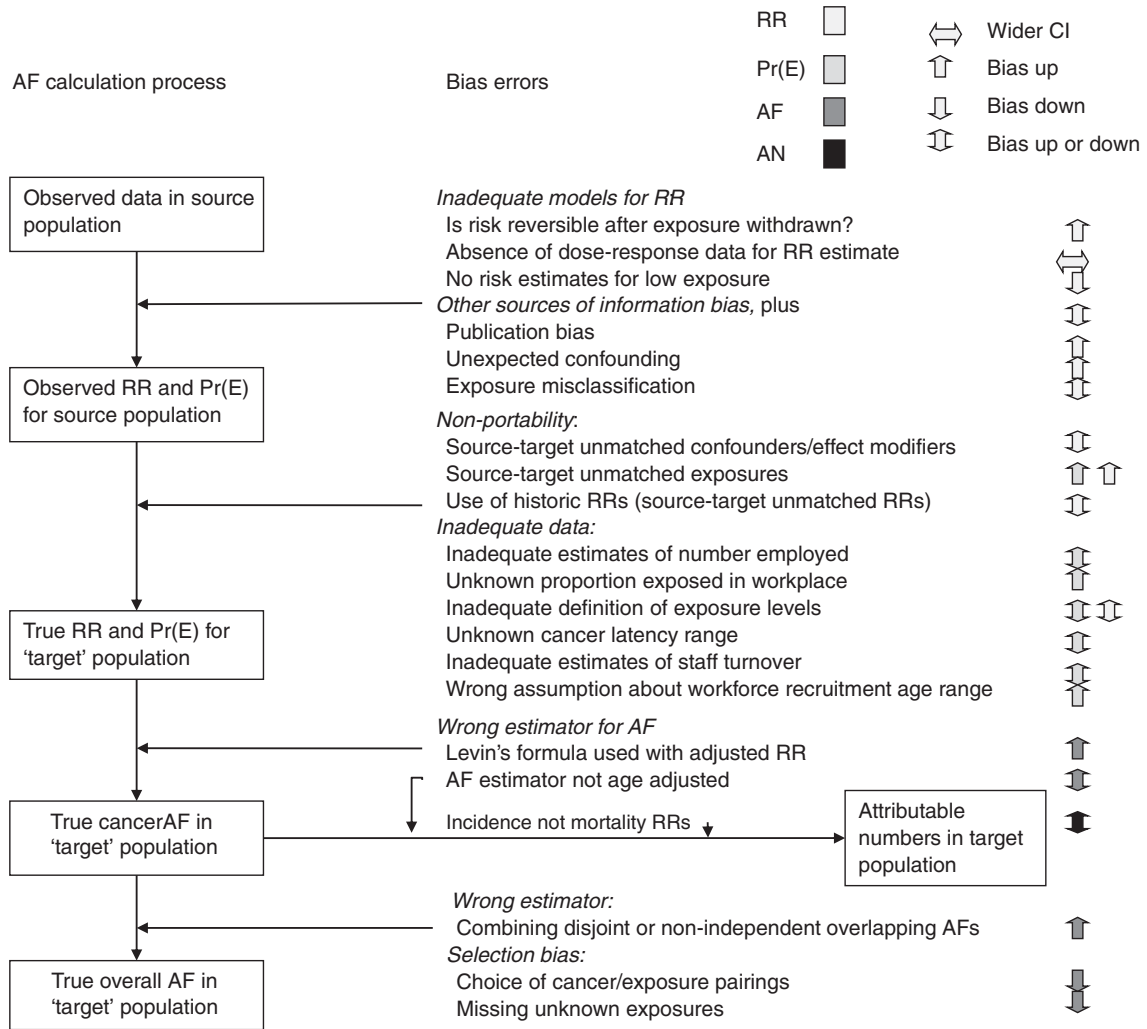
DISCUSSION

A number of assumptions have been made for this study. For the standard REPs, we have assumed a latency period of 10–50 years for solid cancers and for haematopoietic malignancies 0–20 years is assumed. There are limited data available on length of cancer latency, and it has been assumed that latency for cancer due to occupational exposure is the same as that due to all exposures in general.
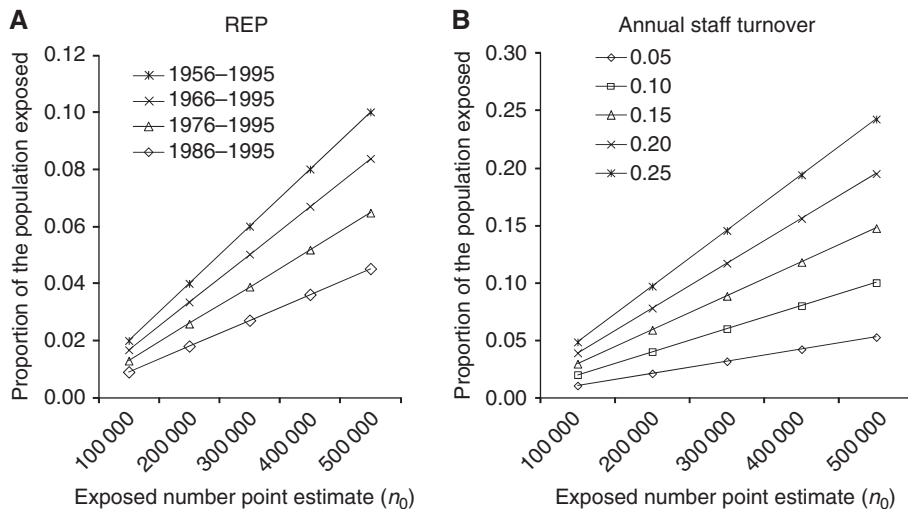
For estimating the number of people ever exposed, it is assumed that there was an even distribution of ages across the cohort of exposed workers in the first year of the REP, and that recruitment to the cohort through turnover was into the youngest age range (15–24) only. It would be possible to adjust the AF estimator to take into account the ages the exposed workers have reached by the target year as a result of making these assumptions, and therefore to take into account their differential (age-related) cancer rates. The effect of making these age adjustments would be to reduce attributable numbers as the workers at risk have a younger age profile that the general population of workers in the REP. However, this would be offset by the higher estimates of attributable cancers that result from more exposed workers surviving up to the target year and hence at risk for a long latency cancer. The effect is illustrated in Figure 3 for a range of possible recruitment age bands; it is noticeable that there is very little difference between using our (15–24) age recruitment assumption without age adjustment of the AF estimator and using older age recruitment assumptions with age adjustment.

It has also been assumed that survival as a result of occupational exposures was similar to survival as a result of non-occupational exposures, as incidence and mortality RRs have been used interchangeably.

There is also simplicity in our approach to estimating the numbers and proportions of the population 'ever exposed across the REP'. Firstly, common assumptions regarding the REP and age structure of the workforce have been used. Secondly, a common set of standardised factors has been used to account for historical changes in GB employment patterns and staff turnovers by industrial sector and for men and women. These estimates are realistic and information based (principally Labour Force Survey data). Thirdly, incorporating survival to the target year (2005) was also important to make realistic estimates of the number of cancers currently still occurring for exposures up to 50 years in the past.

RR	☐	⟷	Wider CI
Pr(E)	☐	⇧	Bias up
AF	▨	⇩	Bias down
AN	■	⇳	Bias up or down

AF calculation process

Bias errors

Observed data in source population

Inadequate models for RR
  Is risk reversible after exposure withdrawn? ⇧
  Absence of dose-response data for RR estimate ⟷
  No risk estimates for low exposure ⇩
Other sources of information bias, plus
  Publication bias ⇳
  Unexpected confounding ⇳
  Exposure misclassification ⇳

Observed RR and Pr(E) for source population

Non-portability:
  Source-target unmatched confounders/effect modifiers ⇳
  Source-target unmatched exposures ⇧ ⇧
  Use of historic RRs (source-target unmatched RRs) ⇳
Inadequate data:
  Inadequate estimates of number employed ⇳
  Unknown proportion exposed in workplace ⇧
  Inadequate definition of exposure levels ⇧ ⇳
  Unknown cancer latency range ⇧
  Inadequate estimates of staff turnover ⇧
  Wrong assumption about workforce recruitment age range ⇳

True RR and Pr(E) for 'target' population

Wrong estimator for AF
  Levin's formula used with adjusted RR ⇧
  AF estimator not age adjusted ⇳

True cancerAF in 'target' population

Incidence not mortality RRs → Attributable numbers in target population ◼

Wrong estimator:
  Combining disjoint or non-independent overlapping AFs ⇧
Selection bias:
  Choice of cancer/exposure pairings ⇩
  Missing unknown exposures ⇩

True overall AF in 'target' population

Figure 4 Sources of bias in the calculation of an overall attributable fraction.

Figure 5 (**A**) Effect of change in length of the solid tumour risk exposure period and shortened latencies on proportion of the population exposed at different levels of exposed number point estimates. (**B**) Effect of change in annual staff turnover on the proportion of the population exposed at different levels of exposed number point estimates.

Different results were obtained for the combined AF depending on whether it was based on mortality or on incidence data. The difference depended on the ratios of deaths to registrations in the cancers that were included (i.e., on survival rates). A combined AF based on attributable deaths would have under-weighted the contribution of non-fatal cancers, such as non-melanoma skin cancer, which is rarely fatal but occurs in relatively large numbers. On the other hand, a combined AF based on incidence would have given all cancers equal weighting so that the large numbers of non-melanoma skin cancers would have made a disproportionately large contribution. This indicates the value of using such estimates as Disability-Adjusted Life Years in conjunction with attributable numbers to estimate an overall AF weighted according to the relative severity of the cancer outcome (especially as survival improves).

Figure 4 outlines the main sources of bias and uncertainty and where they occur in the estimation process. Work is underway to estimate credibility intervals that take into account these sources of bias.

The most influential sources of underestimation of the overall AF are likely to be (1) the choice of exposures we have made and (2) the existence of as yet unidentified workplace carcinogens. Only occupational carcinogens classified by the end of 2008 by IARC as 1 and 2A carcinogens have been included. Other cancer sites, for example, for asbestos have been classified as group 1 or 2A since then, and the review process for newly emerging carcinogens is ongoing (IARC, 2012). These sources of bias are, however, not thought likely to affect the relative rankings of the carcinogens.

For AF estimates for individual carcinogens, we consider the most important sources of bias to be as follows: (1) assumptions about cancer latency and workplace staff turnover and their effect on the estimates of the number of people ever exposed; (2) the absence of estimates of the proportion in the workplace exposed (i.e., where CAREX estimates are not available) for some more recently identified carcinogenic exposures or exposures characterised by occupation only; and (3) the general scarcity of exposure-level measurement data and dose–response RR estimates, leading to the adoption of judgemental allocation of industrial sectors to 'higher' and 'lower' exposure levels and matching to appropriate RR estimates. The lack of individual exposure data is a well-recognised problem; there are few good biomarkers and little monitoring on an individual level for most of the exposures considered in this study.

Of these, the first, although possibly substantial, source of bias will generally affect all the exposures equally, and thus will not affect rank orders. Figure 5 gives an indication of how change in the assumed length of the REP affects estimates of the proportions of the population exposed (Pr(E) in equation (4) above). For a prevalent estimate of 100 000 individuals exposed (n_0 in equation (2), from Labour Force Survey or CAREX data for example), Pr(E) increases by about 0.4% in absolute value for each additional 10 years added to the REP. For $n_0 = 500 000$, the rate of increase is under 2% per additional 10 years. However, this means that if the solid tumour REP (1956–1995) has been overestimated by 10 years (i.e., it should have been 1966–1995) Pr(E) will be overestimated by 19% of the true value, or if the true REP is 1976–1995 Pr(E) will be overestimated by 54% and 125% for a true REP of 1986–1995. These percentage errors are constant across the prevalence estimates (n_0). This will have a substantial impact on the estimate of AF.

For the solid tumour REP, proportions of the population exposed increase approximately in proportion to the increase in turnover, illustrated in Figure 5B for estimates of staff turnover from 5 to 25% a year.

There is no good comprehensive British source of the proportion in the workplace exposed to carcinogenic agents such as a job exposure matrix or a comprehensive national exposure database. Therefore, where there are no CAREX estimates (for
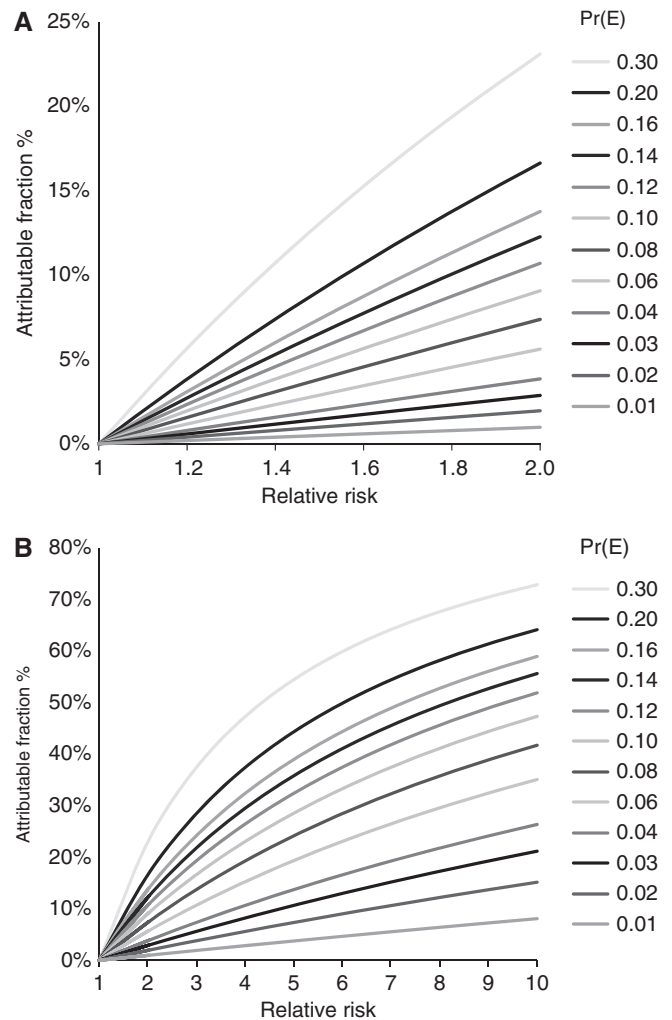
Figure 6 Increase in the attributable fraction for a unit increase in relative risk by proportion of the population exposed (Pr(E)). (**A**) Changes in relative risk from 1 to 2. (**B**) Changes in relative risk from 1 to 10.

mineral oils, TCDD (2,3,7,8-tetrachlorodibenzodioxin-*p*-dioxins), non-arsenical pesticides, welders, painters, hairdressers and barbers, and shift work), we have assumed that all those recorded by occupation are exposed, although these are then matched to appropriate occupation-based rather than agent-based risk estimates. This absence of estimates of the proportion exposed may upwardly bias the AF estimates for exposures. If, for example, only half rather than all of those identified as painters are actually exposed to the unidentified carcinogen(s), then proportions exposed will be halved.

There is an absence of good epidemiological studies for some agents, particularly an absence of dose–response data (often due to a paucity of good exposure measurement data) and of estimates of risk at low levels of exposure. The use of categorical allocation of industrial sectors to higher or lower levels may cause under- or overestimation. In instances where 'lower'-exposed groups have been allocated a zero excess risk, that is, RR = 1, underestimation is probable (e.g., lung cancer due to low-level beryllium, cadmium, chromium, inorganic lead, strong inorganic acid mists exposure; brain and stomach cancers due to low-level inorganic lead exposure; nasopharyngeal and sinonasal cancers due to low-level formaldehyde exposure). When there were insufficient dose–response data to obtain a low-exposure RR, we estimated this based on all available data within the project. In instances where

this resulted in a fairly high RR estimate, overestimation may be a possibility. Examples of this included oesophageal and cervical cancers due to low-level tetrachloroethylene exposure; sinonasal cancers due to low-level wood dust, chromium and mineral oil exposure; and liver cancers due to low-level vinyl chloride monomer exposure (Van Tongeren *et al*, 2012).

Figures 6A and B show, for the range of population proportions exposed and the range of RRs encountered in this study, how a unit change of RR affects the estimate of AF. For example, if the RR for breast cancer due to shift work was 0.1 units lower than the RR used (1.41 instead of 1.51), the AF is reduced by 20% (proportion of the population exposed = 0.09) – Figure 6A. If the zero excess risk allocated for lung cancer due to low exposure to inorganic lead, cadmium and chromium respectively is increased by 0.1 units (RR = 1.1), the AFs for men increase by 180%, 170% and 50%, respectively (for beryllium and strong inorganic acid mists with small numbers exposed at low level, the change is negligible). If the estimated low-exposure RRs for sinonasal cancer due to exposure to chromium, mineral oils and wood dust respectively are decreased by 1 unit (from RR = 3.42 to 2.42 (Figure 6B), 1.85 to 1 and 2.05 to 1.05 respectively; the proportions of the populations exposed are 0.02, 0.30 and 0.09, respectively), AFs for men decrease by 10% and 40% for chromium and mineral oils, but by only 4% for wood dust (as relatively few are classified as low exposed).

Ongoing work has shown that the bias from using Levin's approach with RRs adjusted for confounding factors (Greenland, 1984; Rockhill *et al*, 1998) is close to the ratio of adjusted to unadjusted relative risk and generally likely to be small, and that the use of equation (6) to combine AFs across exposures minimises unavoidable bias.

CONCLUSIONS

In summary, the methodology developed here has allowed us to estimate the overall size of the burden of occupational cancer in Britain, and to assess the relative contributions to that burden from different workplace carcinogens and occupations that have been associated with increased risk of cancer, and from different industries (Hutchings and Rushton, 2012). This allows prioritisation of interventions to address workplace cancer. The methods are robust with respect to the study's objectives, and are adaptable to the limited availability of good exposure data in particular. In instances where better exposure data and dose–response risk estimates are available, the methods can easily be adapted to use these data for more sophisticated modelling. Assumptions and generalisations have been made to enable all of Britain's occupational carcinogens to be addressed on a consistent basis, but the methods allow any parameters to be replaced so that data from other countries and regions and representing change in the future can be substituted. On the basis of the AF methodology described here, a method has also been developed to estimate the future burden of occupational cancer, which can be used to inform strategies for risk reduction (Hutchings and Rushton, 2011). REPs are projected forward in time to estimate AFs for a series of forecast target years based on past and projected exposure trends and under targeted reduction scenarios.

REFERENCES

BASSA (British Airlines Stewards and Stewardesses Association) (2008) Personal communication

Benichou J (2001) A review of adjusted estimators of attributable risk. *Stat Methods Med Res* 10: 195–216

Breslow NE, Day NE (1989) *Statistical Methods in Cancer Research. Vol. II: The Design and Analysis of Cohort Studies*, p 67. International Agency for Research on Cancer: Lyon, France

Darnton AJ, McElvenny DM, Hodgson JT (2006) Estimating the number of asbestos-related lung cancer deaths in Great Britain from 1980 to 2000. *Ann Occup Hyg* 50: 29–38

DerSimonian R, Laird N (1986) Meta-analysis in clinical trials. *Contr Clin Trials* 7: 177–188

Doll R, Peto R (1981) The causes of cancer: quantitative estimates of avoidable risks of cancer in the United States today. *J Natl Cancer Inst* 66(6): 1191–1308

Goldberg M, Imbernon E, Rolland P, Gilg Soit Ilg A, Saves M, de Quillacq A, Frenay C, Chammings S, Arveux P, Boutin C, Launoy G, Pairon J, Astoul P, Galateau-Salle F, Brochard P (2006) The French National Mesothelioma Surveillance Program. *Occup Environ Med* 63: 390–395

Government Actuary's Department (GAD), Historic Interim Life Tables (2007). http://www.gad.gov.uk/DemographyData/LifeTables/Historic_interim_life_tables.html (files whltgbm8006.xls and whltgbf8006.xls; accessed 6 March 2007)

Greenland S (1984) Bias in methods for deriving standardized morbidity ratio and attributable fraction estimates. *Stat Med* 3: 131–141

Greenland S (2004) Interval estimation by simulation as an alternative to and extension of confidence intervals. *Int J Epidem* 33: 1389–1397

Gregg P, Wadsworth J (2002) Job tenure in Britain, 1975–2000. Is a job for life or just for Christmas? *Ox Bull Econ Stat* 64: 111–134

Howel D, Arblaster L, Swinburne L, Schweiger M, Renvoize E, Hatton P (1997) Routes of asbestos exposure and the development of mesothelioma in an English region. *Occup Environ Med* 54: 403–409

HSE (1998) Occupational exposure to ionising radiation 1990–1996. Analysis of doses reported to the Health and Safety Executive's Central Index of Dose Information. http://www.hse.gov.uk/pubns/cidi2.pdf

HSE (2005) Burden of occupational cancer in Great Britain. Summary report of workshop held on 22 and 23 November 2004. Meeting Report HSL/2005/54. http://www.hse.gov.uk/research/hsl_pdf/2005/hsl0554.pdf

HSE (2007) Burden of occupational cancer in Great Britain. Summary report of cancer epidemiology workshop held on 27th and 28th June 2006. Meeting Report HSL/2007/32. http://www.hse.gov.uk/research/hsl_pdf/2007/hsl0732.pdf

HSE (2011) The burden of occupational cancer in Great Britain: Technical report: Methodology: RR595. Health and Safety Executive. http://www.hse.gov.uk/research/

Hutchings S, Rushton L (2011) Towards risk reduction: predicting the future burden of occupational cancer. *Am J Epidemiol* 173: 1069–1077

Hutchings SJ, Rushton L with the British Occupational Cancer Burden Study Group (2012b) Occupational cancer in Britain: Industry sector results. *Br J Cancer* 107(Suppl 1): S92–S103

IARC monographs (2012) http://monographs.iarc.fr/ENG/Monographs/allmonos90.php (accessed April 2012)

Langner I, Blettner M, Gundestrup M, Storm H, Aspholm R, Auvinen A, Pukkala E, Hammer GP, Zeeb H, Hrafnkelsson J, Rafnsson V, Tulinius H, De Angelis G, Verdecchia A, Haldorsen T, Tveten U, Eliasch H, Hammar N, Linnersjo A (2004) Cosmic radiation and cancer mortality among airline pilots: results from a European cohort study (ESCAPE). *Radiat Environ Biophys* 42: 247–256

Labour Force Survey (2009) Available at: http://www.statistics.gov.uk/

Levin M (1953) The occurrence of lung cancer in man. *Acta Unio Int Contra Cancrum* 9: 531–541

Miettinen O (1974) Proportion of disease caused or prevented by a given exposure, trait or intervention. *Am J Epidemiol* 99: 325–332

Nelson DI, Concha-Barrientos M, Driscoll T, Steenland K, Fingerhut M, Punnett L, Pruss-Ustun A, Leigh J, Corvalan C (2005) The global burden of selected occupational diseases and injury risks: methodology and summary. *Am J Ind Med* 48: 400–418

ONS (2006) Mortality statistics. Series DH2 No. 32, Table 2.2

ONS (2007) Cancer statistics: Registrations. Series MB1 No. 35, Table 1

ONS (2009) Census of Employment. Available at: https://www.nomisweb.co.uk/

Pannett B, Kauppinen T, Toikkanen J, Pedersen J, Young R, Kogevinas M (1998) Occupational exposure to carcinogens in Great Britain in 1990–1993: preliminary results. In *CAREX: International Information System on Occupational Exposure to Carcinogens*. Finnish Institute of Occupational Health: Helsinki. Available at: http://www.ttl.fi/en/chemical_safety/carex/Pages/default.aspx (accessed April 2012)

Rake C, Gilham C, Hatch J, Darnton A, Hodgson J, Peto J (2009) Occupational, domestic and environmental mesothelioma risks in the British population: a case-control study. *Br J Cancer* **100:** 1175–1183

Rockhill B, Newman B, Weinberg C (1998) Use and misuse of population attributable fractions. *Am J Public Health* **88:** 15–19

Rothman KJ, Greenland S (1998) *Modern Epidemiology*, 2nd edn, p 660. Lippincott, Williams and Wilkins: Philadelphia, USA

Siemiatycki J, Richardson L, Straif K, Latreille B, Lakhani R, Campbell S, Rousseau M-C, Boffetta P (2004) Listing occupational carcinogens. *Environ Health Perspec* **112:** 1447–1460

Spirtas R, Heineman E, Bernstein L, Beebe G, Keehn R, Stark A, Harlow B, Benichou J (1994) Malignant mesothelioma: attributable risk of asbestos exposure. *Occup Environ Med* **51:** 804–811

Steenland K, Armstrong B (2006) An overview of methods for calculating the burden of disease due to specific risk factors. *Epidemiology* **17**(5): 512–519

United Nations Scientific Committee on the Effects of Atomic Radiation (2008) *Effects of ionising radiation: UNSCEAR 2006.* Vol. 1, Report to the General Assembly, Annex A

Van Tongeren M, Jimenez AS, Hutchings SJ, MacCalman L, Rushton L, Cherrie JW (2012) Occupational cancer in Britain: Exposure assessment methodology. *Br J Cancer* **107**(Suppl 1): S18–S26

Yates D, Corrin B, Stidolph P, Browne K (1997) Malignant mesothelioma in south east England: clinicopathological experience of 272 cases. *Thorax* **52:** 507–512