

The burden of occupational cancer in Great Britain

Methodology

Prepared by the **Imperial College London**
for the Health and Safety Executive 2007

The burden of occupational cancer in Great Britain

Methodology

Sally Hutchings

Imperial College London

Department of Epidemiology and Public Health

Faculty of Medicine

St Mary's Campus

Norfolk Place

London W2 1PG

The aim of this project was to produce an updated estimate of the current burden of occupational cancer specifically for Great Britain. The primary measure of the burden of cancer used was the attributable fraction (AF), ie the proportion of cases that would not have occurred in the absence of exposure. Data on the risk of the disease due to the exposures of interest, taking into account confounding factors and overlapping exposures, were combined with data on the proportion of the target population exposed over the period in which relevant exposure occurred. Estimation was carried out for carcinogenic agents or exposure circumstances that were classified by the International Agency for Research on Cancer (IARC) as Group 1 or 2A carcinogens with strong or suggestive human evidence. Estimation was carried out for 2004 for mortality and 2003 for cancer incidence for cancer of the bladder, leukaemia, cancer of the lung, mesothelioma, non-melanoma skin cancer (NMSC), and sinonasal cancer.

The proportion of cancer deaths in 2004 attributable to occupation was estimated to be 8.0% in men and 1.5% in women with an overall estimate of 4.9% for men plus women. Estimated numbers of deaths attributable to occupation were 6,259 for men and 1,058 for women giving a total of 7,317. The total number of cancer registrations in 2003 attributable to occupational causes was 13,338 for men plus women. Asbestos contributed the largest numbers of deaths and registrations (mesothelioma and lung cancer), followed by mineral oils (mainly NMSC), solar radiation (NMSC), silica (lung cancer) and diesel engine exhaust (lung and bladder cancer). Large numbers of workers were potentially exposed to several carcinogenic agents over the risk exposure periods, particularly in the construction industry, as farmers or as other agricultural workers, and as workers in manufacture of machinery and other equipment, manufacture of wood products, land transport, metal working, painting, welding and textiles. There are several sources of uncertainty in the estimates, including exclusion of other potential carcinogenic agents, potentially inaccurate or approximate data and methodological issues. On balance, the estimates are likely to be a conservative estimate of the true risk. Future work will address estimation for the remaining cancers that have yet to be examined, together with development of methodology for predicting future estimates of the occupational cancers due to more recent exposures.

This report and the work it describes were funded by the Health and Safety Executive (HSE). Its contents, including any opinions and/or conclusions expressed, are those of the authors alone and do not necessarily reflect HSE policy.

© Crown copyright 2007

First published 2007

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means (electronic, mechanical, photocopying, recording or otherwise) without the prior written permission of the copyright owner.

Applications for reproduction should be made in writing to:
Licensing Division, Her Majesty's Stationery Office,
St Clements House, 2-16 Colegate, Norwich NR3 1BQ
or by e-mail to hmsolicensing@cabinct-office.x.gsi.gov.uk

ACKNOWLEDGEMENTS

We would like to thank Damien McElvenny for initiating the project and Gareth Evans for his management role. Andy Darnton from the HSE was responsible for the work on mesothelioma. The contributions to the project and advice received from many other HSE and HSL staff is gratefully acknowledged. Two workshops were held during the project bringing together experts from the UK and around the world. We would like to thank all those who participated and have continued to give advice and comment on the project.

CONTENTS

Acknowledgements:	iii
1. Introduction	1
2. Alternative approaches to calculating occupation attributable burden	2
3. Key features of the approach adopted	5
3.1 Listing the occupational exposures	7
3.2 Handling the exposures	7
3.3 Background Information on each cancer	9
3.4 Review of the literature and selecting the best sources for relative risks, and matching the RRs to the appropriate exposures	9
3.5 The relevant exposure period	10
3.6 Calculating the attributable fraction	11
3.7 Combining the attributable fractions across exposures	14
3.8 Confidence Interval for the overall attributable fraction	14
3.9 Estimating attributable numbers	15
3.10 Combining the attributable fractions across cancer sites	15
4. Examples of the calculation of AF for cancer/exposure pairings	17
4.1 Miettinen's formula with a study based estimate of proportion exposed: Nasal cancer and Formaldehyde	17
4.2 Levin's formula with an independent estimate of proportion exposed: Nasal cancer and Formaldehyde	18
4.3 Comparing results from Miettinen's and Levin's formulae	19
5. Estimating future burden	22
6. Appendices:	
6.1 Appendix 1: Literature review	24
6.2 Appendix 2: Criteria for choosing a 'best study' for RR estimates	25
6.3 Appendix 3: Decision-making process for selecting cancer/exposure pairings	27
6.4 Appendix 4: Criteria for selecting the RRs to use from the 'best study', and matching to appropriate exposure data	29
6.5 Appendix 5: Defining the relevant exposure period	32
6.6 Appendix 6: Calculating the attributable fraction	34
6.7 Appendix 7: Combining AF across exposures	45
6.8 Appendix 8: Random error confidence intervals	47
6.9 Appendix 9: Alternative approach for asbestos exposure, based on the incidence of mesothelioma	49
7. ANNEXES	52
Annex 1: Calculation of CAREX adjustment factors Numbers and percent of men and women in employment by industry and occupation	53 55
Annex 2: Glossary of terms	56
8. Bibliography	58

1. INTRODUCTION

The aim of this project was to estimate an attributable fraction (AF) of cancer due to occupation in Great Britain. The AF is the proportion of cases that would not have occurred in the absence of occupational exposure. There are several methods of estimating the AF but all depend on knowledge of the risk of the disease due to the exposure of interest and of the proportion of the population exposed.

The purpose of estimating the attributable fraction was:

- (1) To estimate a current overall attributable fraction and attributable number of cancers due to occupation.
- (2) To indicate the relative contributions of the occupational carcinogens to the current total burden of occupational cancer in Great Britain, thereby providing evidence on which to prioritise intervention.
- (3) To estimate future burden resulting from current exposures, in order to indicate where to prioritise future intervention.
- (4) To suggest areas for future data collection, to improve the quality of the estimates.

The current best estimate of the proportion of cancer deaths in Great Britain due to occupational causes was 4%, with an uncertainty range of from 2% to 8%, based on the study of Doll and Peto (1981). This equates to approximately 6,000 deaths per annum (with a range of 3,000 to 12,000).

2. ALTERNATIVE APPROACHES TO CALCULATING THE OCCUPATIONAL ATTRIBUTABLE BURDEN

There are a number of basic approaches that have been used to calculate the occupation attributable cancer burden. These are:

1. For each cancer/exposure pairing, a relative risk estimate is obtained from epidemiological studies, and an estimate of the proportion of the population exposed is obtained from census or national employment data or from a job-exposure matrix (JEM), or from the CARcinogen EXposure (CAREX) database.

Examples of this approach are:

- a) Dreyer et al., (1997), who calculated the AF for cancers due to occupation in the Nordic countries, using a common relevant exposure period of 1970-84 for a target year of 2000 and a common of employment turnover (15% for those employed for over one year.
- b) World Health Organisation (WHO) Global Burden of Disease (GBD) Study;
- c) Steenland et al (2003), for the USA;
- d) Nurminen & Karjalainen (2001), for Finland;
- e) Mathers et al., (2000), for Australia, using GBD methodology;
- f) Driscoll et al., (2004), for New Zealand, who used the AFs calculated by Nurminen & Karjalainen plus Steenland's estimates for lung cancer;
- g) Morabia et al., (1992), for the US, for lung cancer only, using RRs from one large hospital-based case-control study;
- h) Steenland, Loomis et al., (1996), for the USA, for lung cancer only, who combined RR estimates by exposure agent from a review of cohort studies, and matched these to National Institute for Occupational Safety and Health (NIOSH) numbers of workers exposed.

2. The results of a pooled analysis of (usually population-based) case control studies are used for an estimate of relative risk, with internal estimates of Pr(E) or Pr(E|D) obtained from the distribution of exposures amongst the controls, or cases.

Examples of this approach are:

- a) Vineis et al., (1988), for the US, for lung cancer, using a pool of population-based case-control studies;
- b) Gustavsson et al., (2000), for Sweden, lung cancer only, using a single population-based case-control study;
- c) Kogevinas et al., (1998) for Europe, who also quoted method 5 below in a review of methodology suitable to produce an AF for Europe.

2a. Similar to the above, estimates of relative risk are obtained from a search of the literature for the 'best studies' (population-based case-control studies) for an occupational cancer, and an AF is calculated from each of these studies using internal estimates of Pr(E) (the proportion of the population exposed) or Pr(E|D) (the proportion of cases exposed (E = exposed, D = case)), to give a range of possible AF estimates. Additional lower and upper estimates are also provided that are based on inclusion of only well-established exposures for the lower estimates, and these plus 'possible' and 'uncertain' exposures for the upper estimates.

An example of this approach is:

- d) Vineis and Simonato (1991), for the USA and Europe

Normally, exposure assessments will not be good enough from population or hospital based case-control studies to estimate a dose-response relationship. To estimate AF, ever/never exposed categories are needed, but it is necessary to have a defined ever/never boundary level. This will normally be less precise from population or hospital based case-control studies, with exposure data generally collected by postal questionnaire or interview rather than expert hygiene assessment where the exposures occurred.

3. Absolute risk measures rather than relative risks are appropriate for a few cancer/exposure pairings for which the AF for a cancer associated with an occupational exposure is thought to be 100%, for example mesothelioma, uniquely caused by exposure to asbestos.

Examples of this approach are:

- a) WHO GBD study (Nelson)
- b) Hodgson & Darnton (2000)

In these cases an age period cohort method is also used to predict future numbers.

4. Proportional Mortality Ratios (PMRs), which can substitute for RRs in the calculation of AF, are available for most cancers by occupation from the Occupational Health Decennial Supplement (1995). The proportion of the population exposed (Pr(E)) in a relevant exposure period would come from national employment data (LFS for occupations), although the reliability of occupation as recorded on the death certificate is known to be problematic, as is matching with other sources of routinely collected data on occupation (Carpenter and Roman, 1999). This method could be used as a 'reality check' on methods (1) and (2) above, where UK studies are not available, and to contribute to confidence intervals. Again an age period cohort method is used to predict future numbers.

4a. Similar to the above, Bouchardy et al., (2002) used cancer registry data alone to estimate occupational odds ratios for Switzerland, using a case-referent method, with the cancer of interest as cases and all other cancers as controls, and job history as recorded by the registry. To extend this approach to obtain an AF, an external source of employment data would be required.

5. Linkage analysis of census and cancer registry data can be used if national databases permit. For example a population based (often a national census-based) cohort can be established with cancer registration or death certificate follow-up.

Examples of this approach are:

- a) Andersen et al., (1999), in Norway, Sweden, Finland and Denmark, linked cancer registry and census data to produce SIRs by occupation, but not AF;
- b) vanLoon et al., (1997) in the Netherlands, for lung cancer only, used a population-based cohort study.

6. The 'Delphic principle' (Morrell et al, 1998) of Doll and Peto, which uses panels of experts to estimate AF.

An example of this approach is that of Landrigan et al., (2002), who examined environmental attributable burden of disease in US children, using panels of experts to arrive at an estimate of AF by a consensus process of meetings and ballots.

7. A proportionate attributable risk approach, which is an extension of the Delphic principle of Doll and Peto, can be used.

Examples of this approach are:

- a) Morrell et al (1998), for Australia, used estimates of attributable percents for all hazardous substances from countries considered to have similar exposure conditions. Estimates from the United States (Markowitz et al., 1989) and Israel (Blanc et al., 1992), plus the Doll and Peto estimates, were applied to numbers of deaths in Australia, using age at death to calculate person-years of life lost (PYLL).
- b) Leigh, Markowitz and Fahs, (1997) also used this method to estimate occupation attributable numbers for the USA, reviewing US-based studies for the established occupational cancers to arrive at an AF range for each cancer, then applying this to cancer mortality for one year, 1992.

8. Another approach is the descriptive study of incident cases. Deschamps et al., (2006) examined all newly diagnosed first cancers occurring over three years (1995-98) in a single county in France, in those over 16 excluding housewives or permanently unemployed, collecting job histories. They considered 13 well-established cancer/exposure pairings only as occupational candidate cancers, and assigned cases as occupational if the individual was exposed for at least 5 weeks per year for a minimum three (not necessarily consecutive) years to the specific occupational exposure. They then calculated the proportion that were occupational by taking these as a percentage of all the cancer cases in the study. The authors also give highest and lowest AF estimates taking the highest and lowest percentages published internationally. This is a straightforward approach, but a problem with this study is that the authors had to consider the distribution of occupations in their cancer patients (in the single French county) to be the same as in the general population (which was supported by a comparison with census data). Excluding non-working cases would also bias the results.

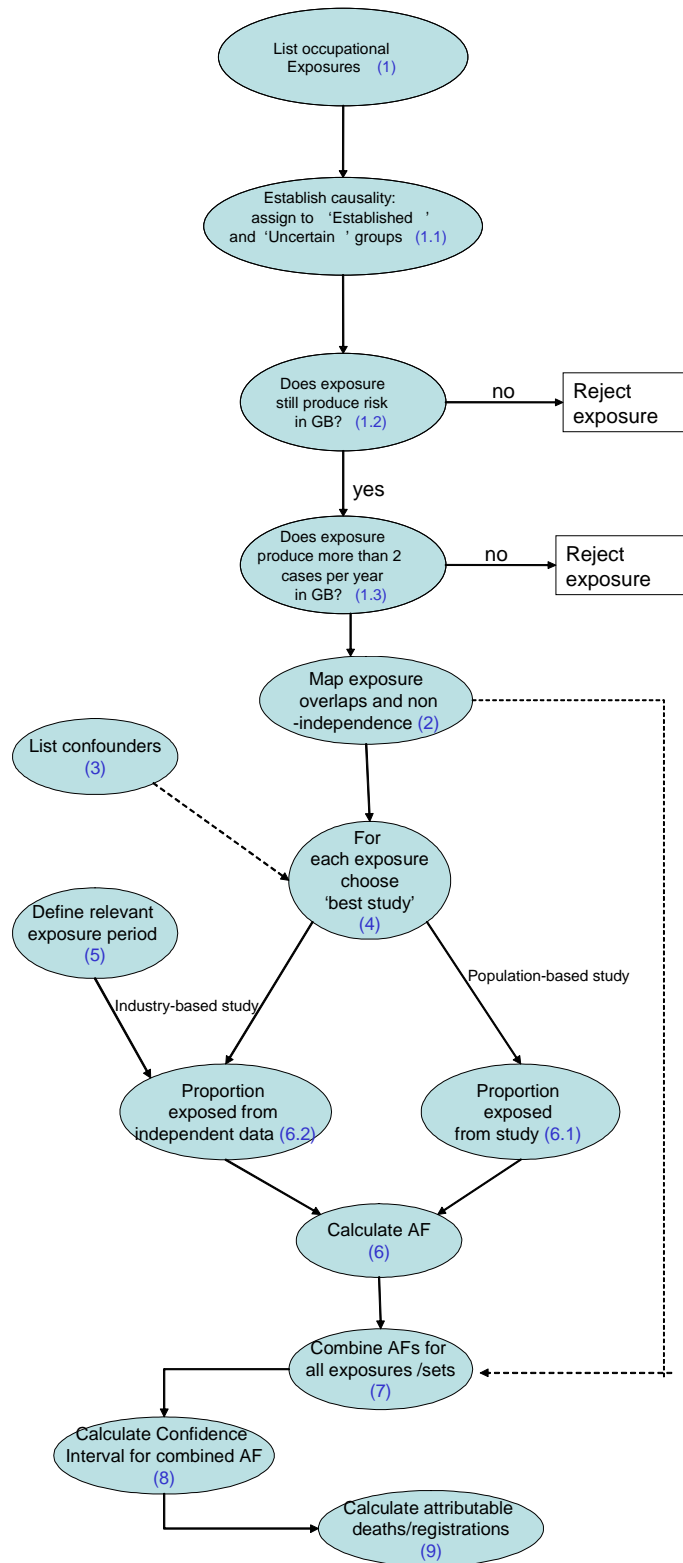
The first two of these approaches have been adapted for this study. Elements of the 'Delphic Principle' also appear, in the choice of 'best study' by a panel of experts from which a relative risk (RR) estimate is extracted for each cancer/exposure pairing. The first approach, and Levin's equation for the AF (see section 3.6), is used where the RR is obtained from an industry based cohort study, or a review or meta-analysis, in which case national estimates of numbers exposed are used. In practice most estimates to date have been obtained using this first approach. The second approach, and Miettinen's equation for the AF, is used where a population-based study is the source of the RR, from a population comparable in the type and distribution of exposures and confounding factors to that in Great Britain (so that the risk estimates are 'portable', see Appendix 2). Where there has been a good population-based study covering a cancer/exposure pairing for which an AF has been calculated using method 1, this second approach based on entirely separate data has been used to support the findings of the first. It will be used also in the estimation of 'credibility limits' which will cover bias as well as random error in the forthcoming sensitivity analysis of our results. The third approach above was used for the nearly unique position of mesothelioma caused by exposure to asbestos.

3. KEY FEATURES OF THE APPROACH ADOPTED

The overall attributable fraction is calculated on a ‘cancer by cancer’ basis. Relative risks are obtained from epidemiological studies considered the most appropriate currently available. The choice of these ‘best studies’ then determines the way in which AF is calculated.

A diagram of the calculation process for the attributable fraction is shown in Figure 1. The numbers in blue in Figure 1 refer to sub-sections in the text, which follows. Detailed methodology and the derivations of factors used in the calculations are in the appendices. A glossary of terms used in this report and accompanying papers is in Annex 2.

Figure 1: Calculation process for the attributable fraction



3.1 *Listing the occupational exposures*

For each cancer, major sources were reviewed to identify the relevant occupational exposures, plus confounders, and any information on latency and trends in mortality, incidence and survival (see Section 3.3). In practice the paper of Siemiatycki et al., (2004) was used as the primary source for identifying the exposures. Exposures needed to pass three tests for inclusion in the analysis, the most important of which is whether the exposure was judged to be causal. Substances or occupations in IARC groups 1 and 2A, and, rarely, group 2B, with ‘strong’ evidence for carcinogenicity in humans as judged by Siemiatycki et al., (2004) for the cancer site being examined were allocated to an ‘Established’ group of exposures, and those with ‘suggestive’ evidence for carcinogenicity were allocated to an ‘Uncertain’ group. See Appendix 3 for precise details.

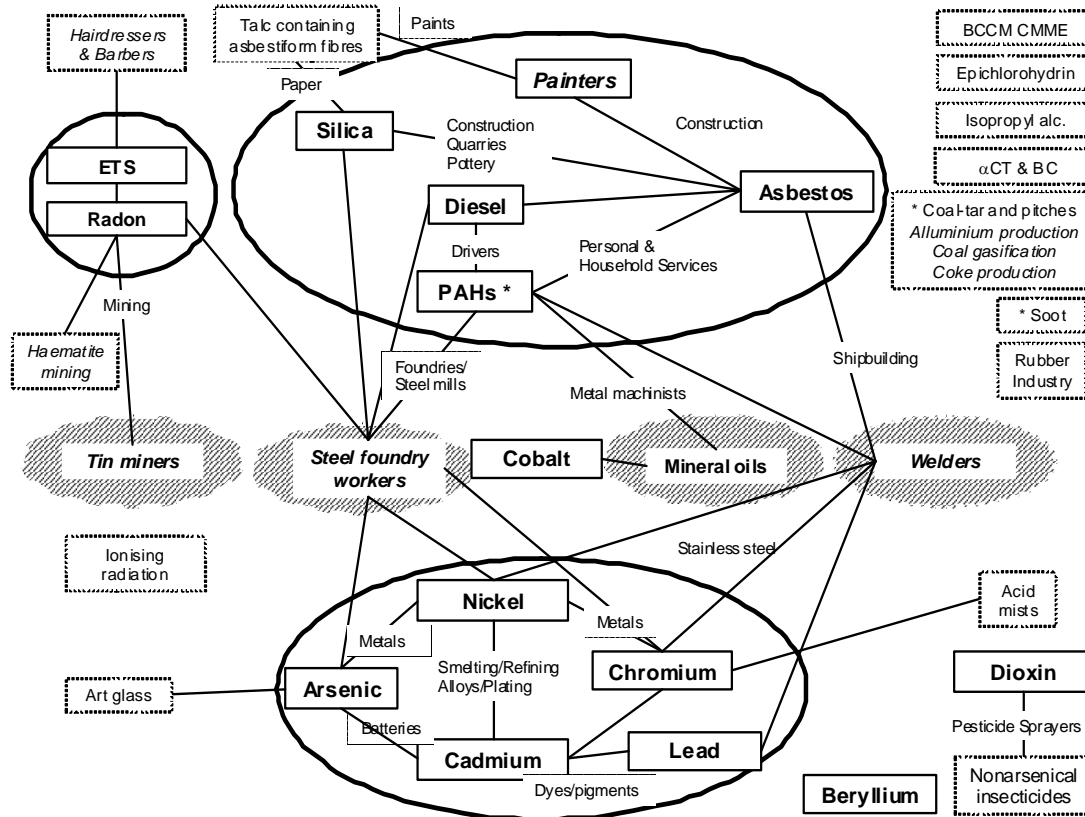
For each exposure, a list of industry and occupation settings (the ‘exposure scenarios’) is also drawn up. In practice this was taken from Siemiatycki et al., (2004).

Due to resource limitations, only the well-established cancer/exposure pairings were tackled initially, that is mainly those in IARC groups 1 and 2A. Relative risks and proportions exposed for the ‘established’ group were easier to obtain and more reliable. However, it was acknowledged that where the ‘established’ group consisted of high risks for rare cancers, or for rare exposures, they may have contributed less to overall occupational morbidity than exposures that were widespread and as yet less well controlled, or which were associated with common cancers, even though the relative risks for this ‘uncertain’ group may have been low.

3.2 *Handling the exposures*

An ‘exposure map’ was drawn up. This is a diagram of how the exposures overlap in the working population, and could interact with one another in causing disease. An example of such a map, for lung cancer, is given in Figure 2. It illustrates the potential for double counting of the exposed population to occur when an overall AF is calculated, and facilitates strategies to avoid this. For a given cancer, the map entries consist of either an agent (or group of agents such as PAHs), or an exposure scenario (i.e. an industry or occupation in which such exposure occurs). Agents are in plain type, exposure scenarios in italics. Lines joining boxes then indicate where overlap would occur were all the entries in the map simply considered separately – for example, if Painters and Asbestos were considered separately overlap would occur in construction (these exposure scenarios are indicated in the smaller print). For substances and occupations shown in grey a separate AF was not estimated, as these exposure scenarios were included with another exposure. More explanation of the content of this map is given in Appendix 7.

Figure 2: Example of an exposure map, for lung cancer



The contribution towards attributable fraction of these overlapping exposures could not (correctly) be estimated separately. These exposures should ideally be treated as a set, and a ‘best epidemiological study’ sought that has relative risk estimates for the set as a whole, or at least that takes account of the confounding effects and interactions within the set. In practice, however, such a study is rarely available. An alternative approach, for more closely defined industry-based exposures, was to partition the proportion of the population exposed between the overlapping exposures, assigning exposed numbers to the ‘highest risk’ or dominant exposure and excluding these workers from ‘lower risk’ estimates. Then the ‘best study’ for the RR estimate was sought for the single exposure. The choice between these two approaches depended on what studies were available.

Single exposures that did not overlap were treated separately when seeking a ‘best study’ for the RR estimate. Single exposures that overlapped but were known to act independently from one another were also treated separately.

The retained exposures were allocated to two groups as described above, the ‘established’ occupational carcinogens and an ‘uncertain’ exposure group. Two attributable fractions and suitable confidence limits (see Appendix 8) were calculated for each cancer/exposure pairing, for the ‘established’ exposures for each cancer and also for exposures in the ‘established’ plus ‘uncertain’ groups combined. The estimate for the ‘established plus uncertain’ groups were regarded as an upper limit for the attributable fraction.

For each cancer/exposure pairing, the approach in general was to partition the exposed population between 'higher exposed' and 'lower' and/or 'background exposure' groups, and apply an appropriate relative risk from the source study or studies to each group separately. Ideally an exposure level threshold was identified that separated the 'higher' and 'lower' exposure groups, but in practice the groups usually were defined by exposure scenarios.

Where no risk estimate data was available from the epidemiological literature, PMRs were available by occupation for most cancers from the Occupational Health Decennial Supplement (OPCS, 1995). These were used to identify additional exposures for the 'uncertain' group and provided estimates of relative risk for the 'upper limit' AF calculation for these groups.

3.3 *Background Information on each cancer*

For each cancer the major non-occupational confounders were listed, and information sought on latency and survival. This information came from the initial review of major sources and from the search of the literature for 'best studies' for each cancer and exposure (see Section 3.4). When relative risks (Section 3.4 below) were chosen preference was given to estimates adjusted for the effects of known confounding factors, and for the effects of interaction between the exposures of interest and these adjustment factors. This was particularly the case for 'ever/never' exposure estimates. For the 'good survival' cancers, RR estimates for incidence rather than mortality had precedence, and the results for registrations as well as deaths were presented, and used in combined calculations.

3.4 *Review of the literature and selecting the best sources for relative risks, and matching the RRs to the appropriate exposures*

The literature was searched for candidate papers for the choice of the best sources, using standard search criteria (see Appendix 1). Information on which to base the choice of best source was entered into a spreadsheet. During the review process all occupational carcinogens were listed, with their associated industry or occupation settings, to support and confirm the initial list (Section 3.1).

The project team then held a review meeting to select the 'best studies'. Papers and other evidence were examined 'exposure by exposure' and standard selection criteria applied, which took into account whether a study represented conditions in GB, and were portable and of good quality (see Appendix 2). Broad-based combined studies (pooled studies and meta-analyses) were usually the first choice, with UK studies brought in if they were appropriate (but had otherwise been excluded). Generally a best source was identified for each exposure scenario (industrial setting or occupation), but studies covering multiple exposures, especially those known to overlap in the working population were also favoured. At this stage exposures that did not contribute significantly to the study were excluded by applying a standard decision-making procedure (outlined in Appendix 3), and the reasons for exclusion were recorded. Apart from the requirement that causality was established, for an attributable fraction to be calculated for a cancer/exposure pairing, the exposure scenario needed to be present in GB during the relevant exposure period for the specific cancer. There also had to be sufficient numbers exposed to produce more than two cases per year.

The relative risks (RRs) that were used in the calculation of AF were identified from the chosen studies, following the criteria set out in Appendix 4. Usually the RR was the study author's preferred 'headline figure'. Occasionally it was necessary to calculate a combined estimate where one was not offered in the chosen review study. An inverse variance weighted average was then calculated if the results passed a

standard test for homogeneity, otherwise a suitable estimate for a random effects model was used (see Appendix A4.1). Separate estimates were needed for the 'higher' and 'lower' level exposed groups. An estimate for the 'higher' exposed group was usually derived from an industry based study or meta-analysis covering the explicit exposure scenario(s). An estimate for the 'lower' exposed came from the same source (if different exposure levels were covered), or from another source such as a broader based population based study. Where no data were available a 'lower exposed' rate was estimated using a linear extrapolation from the 'higher' estimate. This required knowledge of exposure levels at the high and a lower level of exposure, which in practice was rarely available. In a single case where an extrapolation had been used, the higher rate was simply 'halved' (as $1 + (RR_{\text{high}} - 1)/2$). For some exposures, a 'background' exposure level was allocated $RR=1$ (resulting in an AF of zero), so that the exposed numbers remained in the analysis but with a zero AF.

For industry-based studies, data for estimating the proportion of the population exposed was obtained from an external source (see section 3.6 below). In this case the exposure scenarios (the industrial settings or occupations) for which the RRs were calculated were recorded. These were matched by job or industry codes to the external source (see Appendix 4). For population-based studies, data for estimating the proportion of cases exposed (i.e. the numbers of exposed cases contributing to the relative risk estimate, and the total number of cases) was recorded from the RR source study.

3.5 *The relevant exposure period*

The relevant exposure period (REP) is the window of time during which exposure to an occupational carcinogen can result in a cancer being diagnosed or appearing in national mortality or cancer registration records. It is based on cancer latency and is defined in detail in Appendix 5. Distinction should be drawn between the source REP and target REP. The source REP is the period for a source study during which cases contributing to the results of that study (in particular the estimates of relative risk) could have been initiated. It is relevant to the calculation of AF only in determining the portability of the estimate of relative risk between the source and target populations (see Appendix 2). The target REP is the period during which cancers appearing in the national mortality and registration statistics for the target year may have been initiated. It is defined on the estimated maximum and minimum latency for each cancer, and is used in the estimation of the numbers and therefore the proportion of the population that has ever been exposed to the occupational carcinogen being considered.

As there were very little data available on cancer latency, two 'standard' REPs were used for this study, to cover the solid tumours, for which a latency of between 10 and 50 years was assumed, and for the haematopoietic neoplasms such as leukaemia, for which a latency of between 0 and 20 years was assumed. Data specific to these standard REPs used to calculate the attributable fraction are given in Table 1 below.

Table 1: Values of variables used in the calculation of AF specific to the standard relevant exposure periods

<i>Standard REP:</i>	<i>Solid tumours</i>		<i>Haematopoietic neoplasms</i>	
<i>Variable</i>	<i>Men</i>	<i>Women</i>	<i>Men</i>	<i>Women</i>
Latency	10-50 years	10-50 years	0-20 years	0-20 years
Relevant Exposure Period (REP)	1955-1994	1955-1994	1985-2004	1985-2004
'Peak' latency *	35 years	35 years	15 years	15 years
Number of years in REP	40	40	20	20
Total number of those ever of working age during the REP	19.2 million	20.9 million	20.9 million	19.8 million

* 'peak' latency is the latency thought to relate to the highest number of cancer cases leading to current deaths or registrations. However, a uniform distribution of cancer latencies has been assumed across the REP.

3.6 Calculating the attributable fraction

As already noted the method for calculating AF depended on the data source. Data were required for estimates of relative risk and of the proportion of the GB population exposed (or of cases exposed) for each cancer. Three basic approaches were used for the 'established' exposures and any of the 'uncertain' for which data was available. For the first two the 'best epidemiological source' provided an estimate of relative risk, and the type of study determined the equation used to calculate the AF and the way in which overlapping occupational exposures were handled.

If the best study was population-based the estimate of the proportion of cases exposed came directly from the study. A multi-national pool of population-based case-control studies was often in practice a best source. In this case Miettinen's equation (Miettinen, 1974) for calculating the attributable fraction was used:

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

where $\Pr(E|D)$ = proportion of cases exposed (E = exposed, D = case)

If multiple exposures were covered it was sometimes possible to treat overlapping exposures as a 'set' (see Appendix 7). An example of this approach is given for nasal cancer in Section 4.1.

If the best study was industry-based, national data sources combined with the best available estimates of the proportions in the workplace exposed (at relevant levels) provided the estimate of the exposed population. CAREX was often the best and only source for the proportion of the population exposed. Exposed numbers from CAREX, or national employment sources, were allocated to a 'higher' exposure group (based on the industries and occupations covered in the studies), and the others to a 'lower' exposure group. CAREX exposed numbers were split between men and women according to an estimate of relative proportions based on numbers in the appropriate occupation by industry codes from the 1991 Census (details are in Appendix 6 Section A6.5.2).

Adjustment factors were applied to the CAREX data to take account of the change in numbers employed in the primary and manufacturing industry and service sectors in GB between the late 1970s and early 1990s. The factors were estimated from LFS employment data, and were grouped by main industry sector, and for men and women separately (see Appendix 6 Section A6.3). Employment turnover (TO) factors were applied to the point estimates of numbers exposed, which were specific to the same grouped main industry sectors as the CAREX adjustment factors. The turnover factors were estimated using LFS data for the length of time with current employer, which was available by length of employment category (see Appendix 6 section A6.4). The estimates used were for those employed for at least one year. The numbers

‘ever exposed’ therefore depended on the timing and length of the ‘relevant exposure period’ described above. A ‘turnover equation’ was used to calculate numbers ‘ever exposed’ during the REP:

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

where n_0 = numbers employed in middle year of the REP
 TO = staff turnover per year
 t = number of years in the REP

Adapted to include survivors to the target year only, the equation (in a simplified form, allowing for entry into the cohort at age 15 only) becomes:

$$N_{e(\text{REP})} = \sum_{i=65}^{100} \{l_i * n_0 / (R - 15)\} + \sum_{j=25}^{64} \{l_j * n_0 * \text{TO}\}$$

where l_i = proportion of 15-year-olds surviving to age i , summations for REP of 1955 to 1994
 R = retirement age

Life table estimates of the proportions of the general population surviving to the target year (2004) were used to adjust the estimates of numbers ever exposed obtained using the turnover factors, so that only the ever exposed cohort members surviving to the target year would be counted. This part of the estimation process assumed that there was an even distribution of ages across the exposed cohort in its first year (1955 for the ‘solid tumour’ example REP), and that recruitment to the cohort was across the age range of 15 to 24 only.

More detail on how numbers ‘ever exposed’ were calculated in the case of the industry-based studies is outlined in Appendix 6, including the full form of the turnover equation allowing entry into the cohort at multiple ages. The CAREX adjustment factors and turnover factors (TO) used are listed in Table 2. In the last column of Table 2 turnover factors (OT) similar to those used in the Global Burden of Disease methodology (Nelson et al., 2005), are listed and were based on our estimates of length of REP, employment turnover factors and GB life expectancy. The formula used to estimate these factors is given in Appendix 6 (Section 6.4),

Table 2: CAREX adjustment and employment turnover factors used in the calculation of AF

	Main Industry Sector		CAREX adjustment factor	TO	OT equivalent in Global Burden methodology, based on a 40 year REP and GB life expectancy tables
Men	A,B	Agriculture, hunting and forestry; fishing	1	0.09	3
	C,D,E	Mining and quarrying, electricity, gas and water; manufacturing industry	1.4	0.09	4
	F	Construction	1	0.13	5
	G-Q	Service industries	0.9	0.11	4
		Total	1	0.10	4
Women	A,B	Agriculture, hunting and forestry; fishing	0.75	0.10	4
	C,D,E	Mining and quarrying, electricity, gas and water; manufacturing industry	1.5	0.14	6
	F	Construction	0.67	0.16	6
	G-Q	Service industries	0.8	0.15	6
		Total	0.9	0.15	6

Where no CAREX data were available, as is generally the case where an exposure was defined by occupation rather than substance, the LFS was the source most commonly used. Data for 1979 (the earliest year available and closest to the ‘peak’ exposure period in the early to mid 1970s) was used for the solid tumours, and for 1993 was chosen for the shorter latency cancers. As tight as possible a definition of ‘occupation’ was used to minimise overestimation that would result where there was no estimate available of the proportion of workers actually exposed within the target occupation.

To get the proportion of the population exposed, the estimated numbers ‘ever exposed’ during the relevant exposure period were then divided by the total population ever of working age during the same period (given in Table 1 above for the standard REPs). This is based on population estimates by age cohort in the target year.

For some cancer/exposure pairings it was necessary to partition the relevant exposure period between ‘high’ and ‘low’ exposed groups, if an industrial process was known to have changed during the period or controls had been put in place but the exposure has not disappeared. The ‘old’ industry-based relative risks were then applied to the earlier-exposed group, and a lower RR estimate to the more recently exposed group. The method for partitioning the proportion of the population exposed between two periods is detailed in Appendix 6 (Section 6.5.3.).

Levin’s equation was used to calculate the AF in the case of industry-based studies (Levin M, 1953):

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

where Pr(E) = 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 for the ‘higher’ and ‘lower’ exposed groups, and then summed (see Appendix 7).

An example of the use of Levin’s formula and this industry-based approach for a lung cancer exposure is outlined in Section 4 below.

Full details of the methods for calculating AF are outlined in Appendix 6.

For a few cancer/exposure pairings, in particular mesothelioma, which is uniquely associated with asbestos exposure, absolute risk measures rather than relative risks were appropriate. The occupational AF depended only on the degree to which exposures to asbestos were considered to be non-occupational. The method used to calculate AF and attributable numbers for mesothelioma is covered in the Mesothelioma Technical Report.

In the case of lung cancer associated with asbestos exposure, the estimate of attributable numbers and hence of AF was derived directly from the ratio of lung cancer to mesothelioma cases estimated amongst asbestos exposed workers (see Appendix 9). Specialised methods were also used to calculate the AF for lung cancer due to radon exposure, based on estimates of numbers developing lung cancer from exposure to radon in domestic buildings (see Lung Cancer Technical Report, Annex 1).

3.7 *Combining the attributable fractions across exposures*

Attributable fractions cannot in general be summed directly, due to the possibility that workers will have been exposed to more than one occupational carcinogen during their working lifetimes in the REP. There are various strategies to handle overlapping exposures, which are described in Appendix 7, using examples from the analysis for lung cancer based on the exposure map illustrated in Figure 2. The aim of the strategies is to restrict the estimate of the exposed population to eliminate overlaps, in order to be able to sum the AFs directly. However, if it was known that the exposures were independent, and their joint effect on the cancer was multiplicative, the AFs were combined by taking the complement of the product of complements (Appendix 7):

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

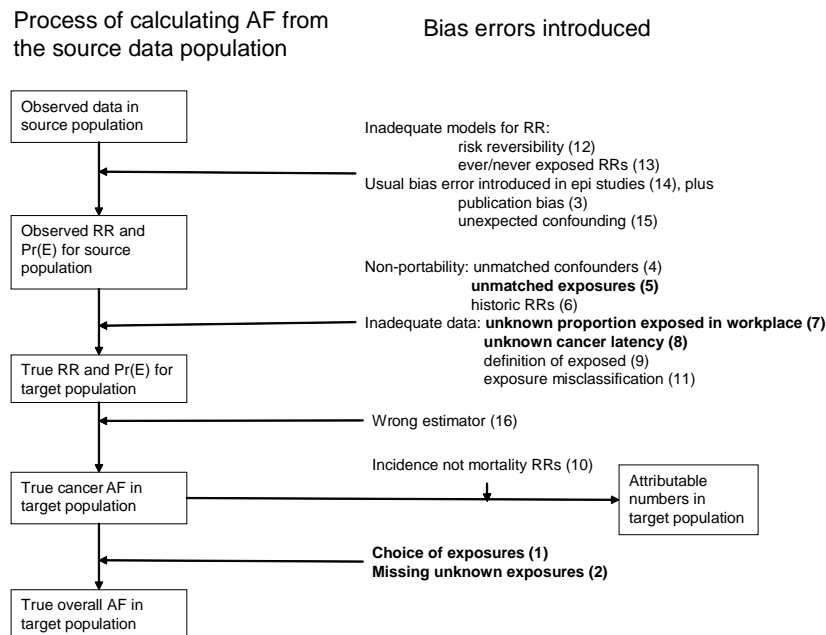
AFs for the established, and for the established plus uncertain, groups of exposures were estimated separately.

3.8 *Confidence interval for the overall attributable fraction*

A 95% random error only confidence interval was calculated for each AF that was based on a single estimate of RR (see Section 3.6 and Appendix 8). No attempt was made to calculate ‘random error only’ confidence intervals for AFs that were summed across exposure levels where more than one estimate of RR had been used, or when the AFs had been combined across exposures.

In future work, Monte Carlo methods (Greenland, 2004) will be used to calculate a ‘credibility interval’ (a confidence interval covering bias as well as random error) for the combined AFs. These limits are able to take into account errors due to bias at all stages of the estimation process, as well as random error in the estimate of relative risk. The details of this work will follow in a subsequent report. An indication of the sources of bias identified in the calculation of AF that need to be accounted for in the ‘credibility interval’ for the AF are given in Figure 3. The sources of greatest bias are in bold (Numbers 1,2,5,7 and 8) in Figure 3.

Figure 3: Sources of bias in the calculation of an overall attributable fraction



3.9 Estimating attributable numbers

Attributable numbers were estimated by multiplying the attributable fraction either by deaths from the cause with the matching ICD code (from DH2 series mortality statistics (ONS, 2005)), or by cancer registrations (from the MB1 series (ONS, 2006)) for the target year. This was the most recent year for which mortality or registration data were available, 2004 for deaths, and 2003 for cancer registrations. For cancers with good or improving survival times, the AFs were applied to registration data as the use of mortality data would have underestimated the burden.

Mortality and registration data were not available for separate histology types; and this was taken into account in seeking ‘source’ study RRs for the AF calculations. Data was also not available for some cancer sub categories (for example some leukaemia sub-types in the Welsh and Scottish data). In this case sub-type estimates were taken from proportions in English data or other sources.

For the purposes of these estimates, it was acceptable to apply attributable fractions calculated from relative risks which were derived from either incident or mortality data, or both as in the case of pooled studies and meta-analyses. In doing this, the assumption was made that survival as a result of occupational exposures was similar to survival as a result of non-occupational exposures. If, for example, survival times were reduced amongst the (perhaps more heavily) occupationally exposed, than amongst the non-occupationally exposed cases, AN would have been underestimated if an incidence AF was applied to mortality data.

3.10 Combining the attributable fractions across cancer sites

To obtain the overall occupation attributable fraction attributable numbers for each cancer were summed, and divided by total cancer numbers. This was acceptable on the assumption that all the cancers being considered as caused by occupational exposure were rare in the GB population. Were they common and a

single exposure was known to produce cancer at more than one site, double-counting an individual's risk would have been an issue, as normally only one primary cancer is recorded in the mortality figures as the underlying cause of death or registration.

Overall attributable numbers and therefore AF can be estimated separately by exposure, or by exposure level, in the same way. Note that attributable numbers (and fractions) summed across cancers by exposure will not normally summed to the overall attributable numbers (and fractions) because of overlapping exposures.

Different results were obtained for the combined AF based on mortality and on incidence data. The difference depended on the ratios of deaths to registrations in the cancers that were included (i.e. on survival rates). An AF based on attributable deaths would have under weighted the contribution of 'good' survival cancers, such as non-melanoma skin cancer (NMSC), which is rarely fatal but occurs in relatively large numbers. On the other hand an AF based on incidence would have given all the cancers equal weighting so that the large numbers of NMSCs would have made a proportionately large contribution. This indicates the value of using such estimates as Disability Adjusted Life Years (DALYs) in conjunction with attributable numbers to estimate an overall AF weighted according to the relative severity of the cancer outcome (especially as survival improves). The role of these estimates will be investigated in the second phase of this study.

It is important to note that if targets are to be set to reduce the incidence of occupational cancer, they must be based on reduced numbers (registrations or deaths) and not on the AF. If cancers due to non-occupational factors are also reduced during the time frame for these targets, the AF will remain static or may well increase even as absolute numbers of cases fall.

4. EXAMPLE OF THE CALCULATION OF AF FOR CANCER / EXPOSURE PAIRINGS

Only one example is given, for nasal cancer and formaldehyde exposure, as both methods were used to estimate an AF for this cancer/exposure pairing. Other examples are given in the technical reports for each cancer.

This example uses alternative methods for estimating AF. The first was based on a RR from a 'pooled population' based case control study using internal study estimates of the proportion exposed (Mannetje et al., 1999), which provided estimates of relative risk for men and for women. The second estimate was based on a UK industry cohort estimate of RR for men only (Coggon et al., 2003), for which data on numbers exposed in GB were taken from CAREX. As women were not represented in the industry cohort, the second estimate also used the RR estimate from the pooled study for the women matched to (estimated) numbers of women exposed from CAREX. This second method was our preferred choice to represent the AF for formaldehyde alone. However, the pooled study estimate for men, using Miettinen's formula for AF, gave a much higher estimate (20%) than the 'industry' based estimate (0%, as the RR=1). Therefore, this pooled study estimate was used to represent the contribution of formaldehyde in the combined exposure AF for nasal cancer, as the 'established' plus 'uncertain' exposures AF was to be used as an 'upper bound' estimate.

This was the only instance to date where a population based study RR and Miettinen's formula was used in the final results. It has generally been the case that where population-based study estimates have been available they have been used to support confidence in the AF that has been obtained from an independent estimate of the proportion exposed in GB.

4.1 Miettinen's formula with a study based estimate of proportion exposed: Nasal cancer and Formaldehyde (Extract from the Nasal Cancer Technical Report)

- (a) **Risk estimate:** In a pooled analysis, by IARC, of 12 case-control studies a non-significant elevated risk, increasing with level of exposure to formaldehyde, was observed among individuals with no or low exposure to wood dust (Luce et al., 2002). Higher and significant ORs were observed among those with moderate and high levels of exposure to wood dust. Luce et al only estimated ORs for adenocarcinomas and for squamous cell carcinomas separately. Mannetje et al. (1999) give ORs from a subset of 8 European studies of Luce's original 12, for sinonasal cancer as a whole, which were adjusted for the effect of other occupational exposures. The ORs were 1.66 (95% CI 1.27–2.17) for men and 0.83 (0.41-1.69) for women for occupational exposure to formaldehyde. These are the ORs used for the AF estimate, with the RR for women set to 1, although the CI values are used in the calculation of a confidence interval for the AF.
- (b) **Numbers exposed:** The proportion estimated from the 'pooled population' based study controls was 34% for men and 17% for women.
- (c) **AF calculation:** Using Miettinen's formula, the AF for men is 20% (95% CI 10% - 29%) and for women is 0 (1.5% - 8%). The results are in Table 4.2.2 below.

4.2 Levin's formula with an independent estimate of proportion exposed: Nasal cancer and Formaldehyde (Extract from the Nasal Cancer Technical Report)

(a) Risk estimate: A cohort of British chemical workers exposed to formaldehyde was established in the early 1980s (Coggon et al., 2003). Five of the six companies involved produced their own formaldehyde on site and either used it to manufacture resins and adhesives, or exported the product as formalin, paraformaldehyde or alcohols. The last company imported formalin to produce resins. The cohort comprised 14,014 men who were followed up to 2000. Between 1941 and 2000 the SMR was 0.87 (95%CI 0.11-3.14), with only two cases observed. Coggon's RR has been used for the lower and background exposure levels for men, taken as RR=1, as a negative AF would result from RR<1 which is not realistic (see below). For the high level exposure group and for women at all exposure levels the RRs from the European case-control study (Mannetje et al, 2000) as for the previous estimate have been used.

(b) Numbers exposed: The numbers of workers exposed to formaldehyde in 1990-1993 according to Carex are given in Table 4.2.1 below. In order to split the Carex exposed numbers between men and women, it is assumed that all the exposed occupations in manufacturing and in construction are in skilled trades, shop floor and transport operatives (SOC major groups 5, 8 and 9), and that the exposed occupations in the service sector are in professional, associate professional and technical and personal and protective service occupations (SOC groups 2, 3 and 6). These data are used to estimate Pr(E) for Levin's calculation of AF (See Annex 1, Table A2).

Table 4.2.1: Numbers of workers exposed to formaldehyde according to Carex in 1990-1993.

Industry	Carex Data 1990-1993		
	Number Exposed	Number in Industry	Exposure Level
Crude petroleum and natural gas production	656	53300	B
Beverage industries	881	88100	B
Manufacture of textiles	4730	182000	L
Manufacture of wearing apparel, except footwear	17992	189500	L
Manufacture of wood and wood and cork products, except furniture	12430	132975	L
Manufacture of furniture and fixture, except primary of metal	39772	144325	L
Manufacture of paper and paper products	722	119050	L
Manufacture of industrial chemicals	1006	130000	L
Manufacture of other chemical products	360	175175	L
Manufacture of plastic products nec	2021	136900	L
Manufacture of glass and glass products	278	43275	L
Manufacture of other non-metallic mineral products	585	70875	L
Iron and steel basic industries	1870	48425	B
Non-ferrous metal basic industries	1254	79325	B
Manufacture of fabricated metal products, except machinery and equipment	535	292200	L
Manufacture of machinery except electrical	760	692275	B
Construction	4511	1753450	B
Education services	122	1455875	B
Research and scientific institutes	176	91100	H
Medical, dental, other health and veterinary services	2796	1435675	H
Recreational and cultural services	74	534600	B

Industry		Carex Data 1990-1993		
		Number Exposed	Number in Industry	Exposure Level
Personal and household services		276	686750	B
Total		93807	8535150	
Main Industry Sector			% Male	
Agriculture, hunting and forestry; fishing	High	0		
	Low	0		
Mining/quarrying, electricity/gas/steam, manufacturing industry	Low	80431	76%	
	Background	5421	76%	
Construction	Low	0		
	Background	4511	99%	
Service industries	High	2972	45%	
	Background	472	45%	

(c) **AF calculation:** Using Levin's formula, based on the above RRs an **AF of 0.02% (0.01% - 0.03%)** was estimated for men and **0 (0 - 0.6%) for women**. These results are summarised in Table 4.2.2 below. This was our preferred estimate for nasal cancer due to formaldehyde exposure (considered independently), although the estimate based on Miettinen's formula was used in the overall nasal AF to reflect our uncertainty about this result.

4.3 Comparing results from Miettinen's and Levin's formulae

To allow a direct comparison with the results from Miettinen's formula, Levin's formula was also used to estimate AF based on the same RRs as in section 4.1 above but the independent estimate of numbers exposed from the data in section 4.2 above (CAREX, undifferentiated by exposure level). **The AF for men was 1.2% (0.5% - 2.0%), and for women 0% (0% - 0.6%).** The difference in men is due to the 20-fold difference in the proportion of the population exposed as estimated from the two different data sources (34% from the study data, 1.8% from the independent CAREX data). This indicates one or more of three possibilities.

- That the European population-based study estimates of numbers exposed are not portable to Great Britain,
- That a higher estimate of RR would be appropriate for the fewer numbers of more highly exposed individuals represented in the independent GB data,
- That there must be serious uncertainty about the reliability of either estimate, due mainly to the difficulty of separating the effects of this exposure from that of wood, leather and textile dust.

In the case of (b), evidence from Coggon et al. (2003), in the UK industry suggested a lower rather than higher RR would be appropriate for GB, as there was no raised risk observed in the industry cohorts.

Table 4.2.2: Results for Nasal Cancer and Formaldehyde Exposure

<i>Occupational exposure</i>		<i>Formaldehyde</i>													
<i>'Best study' for RR estimate</i>	Reference	<i>Mannetje et al (1999)</i>	<i>Mannetje et al (1999)</i>	<i>Coggan et al (2003)</i>						<i>Mannetje et al (1999)</i>	<i>Mannetje et al (1999)</i>				
	Type of study	<i>Pooled population-based case-control studies</i>	<i>Pooled population-based case-control studies</i>	<i>UK industry cohort</i>						<i>Pooled population-based case-control studies</i>	<i>Pooled population-based case-control studies</i>				
	Sex	<i>Male</i>								<i>Female</i>					
	Exposure level		<i>Higher</i>	<i>Lower + Background</i>				TOTAL		<i>Higher</i>	<i>Lower + Background</i>				TOTAL
<i>Independent data:</i>	Industry Sectors	TOTAL	G-Q	C-E	F	G-Q	Total		TOTAL	G-Q	C-E	F	G-Q	Total	
	CAREX numbers exposed		1,337	65,248	4,466	212	69,926	71,263		1,635	20,604	45	260	20,909	22,544
	CAREX adjustment factor		0.9	1.4	1.0	0.9				0.8	1.5	0.7	0.8		
	Annual employment turnover		0.11	0.09	0.13	0.11				0.15	0.14	0.16	0.15		
	Numbers exposed in the REP (1955 - 1994)		5,029	315,606	21,886	799	338,291	343,319		7,827	171,228	192	1,243	174,774	176,209
<i>Study data</i>	Exposed cases	229							15						
	Total cases	451							104						
	Proportion of controls exposed	0.34							0.17						
<i>Proportion of the population exposed</i>			0.0003		0.001		0.018	0.018		0.000		0.000		0.008	0.008
<i>Proportion of cases exposed</i>		0.51							0.14						
<i>Relative risks</i>		1.66	1.66		1		1		0.83	1		1		1	1
<i>Attributable fraction</i>	Levin's		0.00017		0		0	0.0002		0		0		0	0
	'Random error' 95% confidence interval		[0.0001 - 0.0003]		[-0.001 - 0.002]		[-0.016 - 0.036]	[0.0001 - 0.0003]		[0.000 - 0.000]		[0.000 - 0.000]		[-0.005 - 0.006]	[-0.005 - 0.006]
<i>Attributable deaths</i>			0		0		0	0		0		0		0	0
<i>Attributable registrations</i>			0		0		0	0		0		0		0	0

<i>Attributable fraction</i>	Miettinen's	0.20							0						
	'Random error' 95% confidence interval	[0.10 - 0.29]							[-0.15 - 0.08]						
<i>Attributable deaths</i>		14							0						
<i>Attributable registrations</i>		44							0						

5. ESTIMATING FUTURE BURDEN

The above methods can be used with estimates of the proportion of the population exposed based on current levels of exposure and exposed numbers (the REP is shifted forward in time). The key data that will be needed are estimates of the proportions of workers within the industry and occupation categories used for this analysis who are currently exposed at 'higher' and 'lower' levels. A national job exposure matrix would supply this information. It is reasonable to apply existing estimates of relative risk so long as they are applied to the proportions of the population 'known to be exposed' at the levels relevant to these estimates.

6. APPENDICES

Appendix 1: Literature review

Sources for the preliminary review for each cancer included the IARC monographs¹, NIOSH reports² and the US National Toxicology Program³, the ONS DH and MB Series' for mortality and incidence data and the Decennial Supplements on Occupational Mortality.

Searches were made of the PubMed, Web of Science and Medline databases using the following search criteria: "Chemical" + "Cancer" + ("Review" or "Meta-analysis" or "Pooled analysis" or "Occupational"). This was supplemented by references from review papers and textbook bibliographies. Data was extracted from each candidate paper into an EXCEL spreadsheet, which enabled a decision about which paper(s) to choose as the 'best source' for an estimate of relative risk for one or more exposures. The information extracted was placed under the following headings:

- 1) Reference
- 2) Exposure
- 3) Country
- 4) Type of Study: population-based, hospital-based, industry
- 5) Study Size: Cohort, number of cases and controls
Recruitment period
- 6) Exposure Assessment
- 7) Statistic: SMR, RR, OR, etc.
Follow-up period
- 8) Analysis: Basic (including by sex); Adjusted; Dose-Response; Dose-Response, adjusted;
Duration; Year of Hire; Time since first exposure; Other Exposure Metrics
Different jobs/chemicals
Other cancers

¹ <http://monographs.iarc.fr/ENG/Monographs/allmonos90.php>

² <http://www.cdc.gov/niosh/homepage.html>

³ <http://www.cdc.gov/niosh/homepage.html>

Appendix 2: Criteria for choosing a ‘best study’ for RR estimates

The study needed to be broad based and representative of the occupations or industrial exposures that were encountered in Great Britain, and had to cover women as well as men, if this was relevant to the exposure and cancer. Other criteria were those that defined a good epidemiological study. These were a large sample size (so that estimates were robust); an appropriate comparison population; national and large in number for a population-based study; and controlled for the non-occupational confounders identified as important. Adequate exposure assessment was vital, as was a clear and standardised case definition. For example, the ICD codes had to be identified so the results could be applied correctly to national mortality and incidence data.

Portability of the relative risks was important. For example, for a population-based study from which the proportion of the population exposed was to be used, portability between the source (study) population and the target (GB) population was critical. There were three main issues about portability. The most important of these was that the types and levels of exposures matched. This matching was required across populations, so that it would be preferable to use GB based studies or those from countries with a similar industrial heritage and level of development. It was also required across time, so that exposures from the past, in the source relevant exposure period (REP, see Appendix 5), were similar to those in the target REP. It was, therefore, important that the source relevant exposure period was not so long ago that the exposures on which the relative risk estimates were based no longer matched. Of secondary importance was that the distribution of known confounders in the source and target populations matched.

Studies were required for all relevant occupations, chemicals, industries or occupational situations that were identified (the exposure scenarios). The nature of the substance, and of the tasks, jobs and processes needed to be similar to the GB situation.

Large pooled studies or meta-analyses and reviews offering cross-study estimates were likely to be the studies of first choice. The advantages of combined analyses were increased precision due to increased sample size and the opportunity to explore potential confounders, interactions and modifying effects that may explain heterogeneity among studies (Greenland, 1998). A disadvantage of combined analyses was the possible lack of compatibility of data from various studies due to differences in subject recruitment, procedures of data collection, methods of measurement and effects of unmeasured covariates that may differ among studies⁴. Heterogeneity was an important consideration in combined analyses and needed to be accounted for (see Appendix 4).

Population-based studies had the potential advantage of covering multiple exposures and so accounting for overlaps in the working population between these exposures. They were less likely to provide precise definitions of exposure as they are based on ‘self-reported’ occupational histories. These were also generally limited to an ‘ever/never’ exposed dichotomy. Population-based studies were also unsuitable to pick up rarer occupational exposures, unless they were very large. Industry-based studies were likely to provide more precise exposure data analysed by level and duration of exposure, but require comparable independent estimates of the proportion of the GB population exposed.

Odds ratios (ORs) from case-control studies, standardised mortality ratios (SMRs) from cohort studies or proportional mortality ratios (PMRs), were all used as RR estimates in the calculation of AF. In the case of ORs however, the ‘rare disease’ assumption (that the probability of disease was very small) needed to be satisfied. The choice of ‘best study’ did not therefore depend on whether it was a case-control or cohort study (although the quality of exposure assessment may have been an issue here). Whether the study was

⁴ IARC Monograph Preamble, 2006

population-based or industry-based, however, determined the way in which the attributable fraction was calculated.

Where a UK study was available, this was always considered for inclusion. This was achieved for example by combining the results from the UK study with a 'best study' review from which this study was omitted.

Appendix 3: Decision-making process for selecting cancer/exposure pairings

For each cancer site, firstly all known and suspected occupational carcinogens, or exposure scenarios (usually occupations) where the causal agent may be unknown, were listed. The main source for this list was the substances and occupations classified by IARC as group 1 and 2 carcinogens, as tabulated by Siemiatycki et al., (2004). This list was supplemented by information from the HSE Manchester Burden of Cancer workshop (2004)⁵, the EU list and from a priority list of carcinogens drawn up for a carcinogen hygiene profiling exercise developed for HSE's Disease Reduction Programme. The Occupational Health Decennial Supplement was an additional source of information.

For each exposure, a list of occupations and industries in which the exposure was known to have occurred was then drawn up. These were the exposure scenarios. The main source for this was Siemiatycki et al., (2004), specifically Tables 3, 4 and 6 from this paper.

Each exposure listed then had to pass three tests to be retained in the analysis, as follows:

Test 1: Was the exposure causal? From this test, the exposures were allocated to one of two groups based on the strength of evidence for causality. Substances or occupations in IARC group 1 and with 'strong' evidence of carcinogenicity in humans for the relevant cancer site (from Siemiatycki et al., 2004, Tables 3 and 6) were allocated to an 'established' exposure group. Substances or occupations in IARC group 1 or 2A and with 'suggestive' evidence of carcinogenicity in humans for the relevant cancer site (from Siemiatycki et al., 2004, Tables 3, 4 and 6) were allocated to a second tier 'uncertain' group. Additional exposures identified from the other sources were allocated to the 'uncertain' group unless there was 'strong' evidence to the contrary, probably founded in UK experience.

Test 2: Did the exposure scenario (i.e. occupation or industry situation) exist in Great Britain during the target 'relevant exposure period'? This was defined by cancer latency (see below.) For each industry or occupation listed for an exposure, if the industry did not exist in GB, or if processes had been changed by the mid 1950's for solid tumours, or by around 1985 for the shorter latency neoplasms, it was excluded. If the industry had disappeared or the processes had changed by the early 1970's for the solid tumours or by around 1990 for the shorter latency neoplasms, this was accounted for in the analysis. If effective controls were known to have been introduced before this time, but the exposure scenario still existed (or exposure levels had otherwise been substantially reduced), this was noted and accounted for in the analysis.

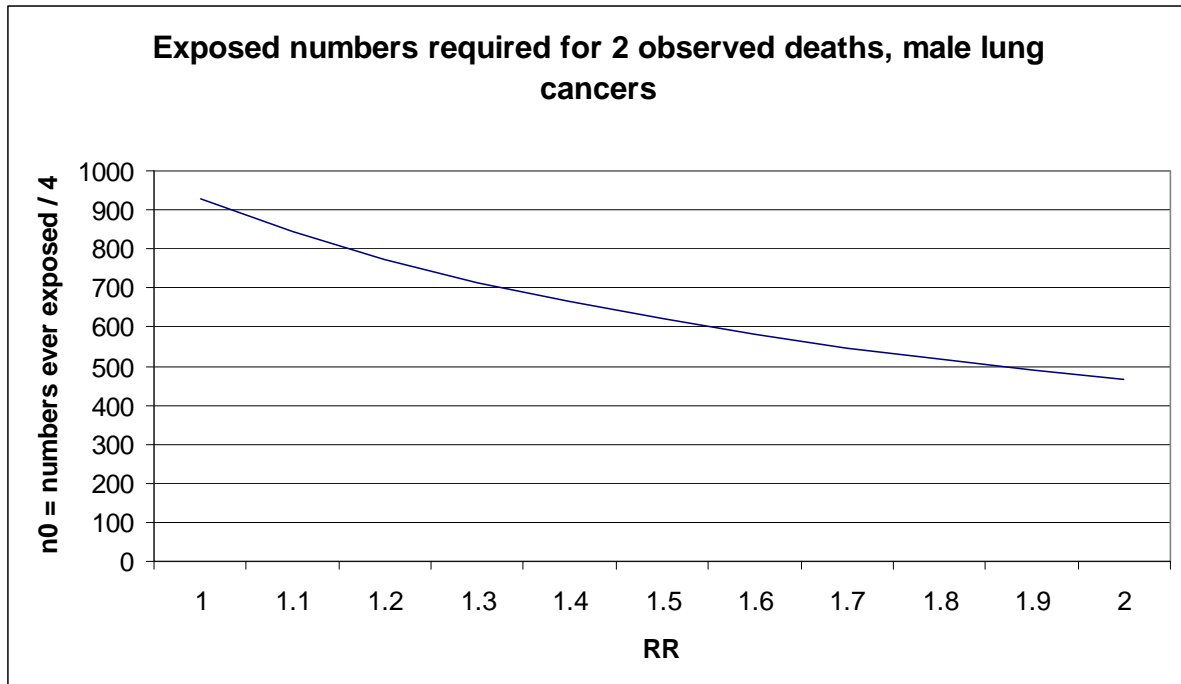
Test 3: Did the exposure scenario (occupation or industry) produce more than two cases per year? Cases would be the number of deaths for short survival cancers, or registrations for those cancers with better treatment outcomes. This equated to a minimum of approximately 1,000 currently exposed workers (given employment turnover rate assumptions) and a relative risk of 1.5 (see Figure 3.1). The CAREX tables of exposures by agent and industry⁶ for GB were the usual source for making this judgement.

Where there was a clear overlap between exposed populations subject to individual exposure scenarios for a single cancer, the exposure scenario with the lower RR estimate was dropped from the calculation of overall AF for that cancer to avoid double counting. It was excluded from the analysis altogether if it was clear from a 'best study' that the estimated raised risk was attributable to another (dominant) exposure scenario.

⁵ GB Burden of Occupational Cancer: Summary Report of Workshop Held on the 22nd and 23rd November 2004 in Manchester: HE/05/03

⁷ CAREX: Industry specific estimates, Great Britain 1990-93. Finnish Institute of Occupational Health.

Figure 3.1



Appendix 4: Criteria for selecting the RRs to use from the ‘best study’, and matching to appropriate exposure data

A key relevant paper, or ‘best study’, was needed for each exposure scenario that had been identified.

Relative risks adjusted for the effects of known confounding (adjustment) factors, and for the effects of interaction between the exposures of interest and adjustment factors, were normally preferred. This was particularly the case for ‘ever/never’ exposure estimates.

For each exposure a minimum level at which disease can be induced was identified explicitly. This usually came from the choice of a cut-off in the chosen ‘best study’ although, particularly for population-based studies relying on self-reported employment history, occupation was sometimes all the information given.

When the ‘best study’ was a review, or meta-analysis, that did not offer a suitable pooled or combined estimate, an average overall estimate (inverse-variance weighted) of the relative risk and its (95%) confidence interval were calculated from the data. This estimate was based on a fixed effects model if the standard test for heterogeneity of the constituent RRs was not significant (at the 95% level), otherwise a random effects model was chosen (see Section A4.1 below).

The exposed population was generally partitioned between ‘higher exposed’ and ‘lower exposure’ groups, and an appropriate relative risk from the source study or studies was applied to each group separately. Ideally an exposure level threshold was identified that separated the ‘high’ and ‘low’ exposure groups, but in practice exposure scenarios usually defined the groups.

For population-based studies a single ‘ever/never’ exposed RR estimate was sometimes all that was available. This may have been applicable only to a ‘low exposure’ group. However if the ‘higher level’ exposure was widespread enough, or the study large enough to identify high exposure cases, RR estimates were taken from the study covering both the ‘high exposure’ and ‘low exposure’ groups. If all the exposure scenarios covered in the study were portable to the GB situation, the overall ‘ever/never’ exposed RR was correctly applied across the combined ‘high’ and ‘low’ exposed groups. If some exposure scenarios were not relevant to the GB situation, RRs were chosen that applied only to the relevant scenarios.

For single industry-based studies the relative risks were normally applied to the ‘highly exposed’ group. The estimates were usually attached to a specific exposure scenario (industry situation or occupation), but estimates by dose level, or dose combined with duration of exposure could be used, thus providing ‘high’ and ‘low’ exposure estimates, depending on what was available from the ‘best study’ source. Otherwise the ‘low exposure’ group RR was estimated from whatever data was available from the ‘best study’ (if it was a review or meta-analysis the RRs from constituent studies covering groups previously allocated as ‘low’ exposed were used), or it came from a separate population-based study.

In the absence of any ‘best study’ data it was sometimes necessary to assume a linear extrapolation between the ‘higher exposed’ risk estimates and never exposed (RR=1) for the ‘lower’ group. For some cancer/exposure pairings, a negligible or ‘background’ exposure group was recognised. In these circumstances an RR=1 was assigned to these groups which resulted in a zero AF, but this allowed recognition in the tabulated results of their zero contribution.

The exposed numbers were partitioned to match the ‘higher’ and ‘lower’ RR estimates, on the basis of job/industry codes. In the case of CAREX data, exposures were coded to the UN ISIC Rev2 (1968) classification, which were matched to SIC68 codes. Exposed numbers for codes that matched the ‘best

study' high exposure scenarios were assigned to the 'higher exposed' group, and other codes to the 'lower exposure' group, so long as the exposure scenario passed the inclusion tests described in Appendix 3 (see Appendix 6.5).

Where there was no CAREX data, national employment estimates were used. Employee counts were needed for industry, or occupation codes, that matched as precisely as possible to the occupation and industries for the relative risks derived from the best study. Data were available from the Census of Employment (CoE) for the period 1971 to 1991, and from its successors (the Annual Employment Survey (AES) and the Annual Business Inquiry (ABI)) for 1991 – 1998 and 1998 – 2004 respectively, or the Labour Force Survey (LFS) from 1979 onwards. The CoE, for industry-based exposure classes, or the LFS, for occupations (earliest available date 1979) were used for exposed counts for 1973 for solid tumour cancers, and the AES and LFS for 1993 for the shorter latency cancers (see Appendices 5 and 6). These counts also were allocated to 'higher' and 'lower' exposure groups on expert judgement, unless there was any data available from HSE (the National Exposure DataBase for example) to base this on.

Clearly there were potential problems in matching RRs, which were based on precisely defined levels of exposure to the (historic) GB working population based on occupation or industry codes without a detailed job/exposure matrix or good exposure assessment information. Expert judgement was required in these circumstances.

A4.1 Calculating a weighted average relative risk and its confidence interval

Occasionally it was necessary to calculate a weighted average across a range of SMRs or ORs where a suitable overall estimate was not offered by a 'best study'. In these circumstances a precision (inverse variance) weighted average was calculated (Rothman, K.J and Greenland, S., 1998). The weights were the inverse of the squared standard error:

$$w = 1/SE^2$$

$$\text{so that } RR_{\text{overall}} = \sum_i (w_i RR_i) / \sum_i w_i \quad \text{where } RR_i = \text{relative risk}; w_i = \text{weight for each estimate } i$$

The standard error of this estimate is given by:

$$s = 1/(\sum_i w_i)^{1/2}$$

For RR estimated by an SMR (Breslow and Day, 1989):

$$SE(\ln(\text{SMR})) = 1/D^{1/2} \quad \text{where } D = \text{observed deaths}$$

$$w = D$$

$$\ln(\text{SMR}_{\text{overall}}) = \sum_i (w_i * \ln(\text{SMR}_i)) / \sum_i w_i$$

$$s = 1/(\sum_i D_i)^{1/2}$$

Therefore

$$\text{SMR}_{\text{overall}} = \exp\{\sum_i (w_i * \ln(\text{SMR}_i)) / \sum_i w_i\}$$

and an approximate 95% confidence interval for the weighted average estimate was given by:

$$CI = \exp(\ln(SMR_{\text{overall}}) \pm 1.96*s)$$

For RR estimated from an OR:

$$SE(\ln(OR)) = (\ln(UCL) - \ln(LCL)) / (2*1.96) \quad \text{where UCL and LCL are the upper and lower 95\% confidence limits for OR}$$

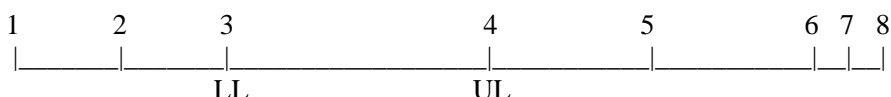
this was continued as for the SMR.

This was the estimate for a fixed effects model. The estimate for a random effects model was calculated using the methods described in DerSimonian and Laird (1986). This was programmed in the AF calculation EXCEL spreadsheets, or obtained using MIX software.

Appendix 5: Defining the relevant exposure period

The ‘relevant exposure period’ (REP) was the period during which exposure to a causal factor resulted in a cancer mortality or registration. This depended on the cancer latency.

Below is an example of an individual case time line for a specific cause of death related to an occupational exposure:



Key:

- 1 age 16
- 2 start of employment in a relevant industry or occupation
- 3 LL = start of REP
- 4 UL = end of REP
- 5 control measures introduced /process closed
- 6 start of risk period (RP)
- 7 death from the specific cause
- 8 end of RP

where for the case (LL) to (7) is the maximum latency (I_{\max})
 (UL) to (7) is the minimum latency (I_{\min})
 In general (LL) to (6) is I_{\max}
 (UL) to (8) is I_{\min}

If points 1 and /or 2 occurred after LL, 2 (or if unavailable then 1) replaced LL as the start of the individual’s REP.

Where an exposure occurred over a defined period of time only (for example prior to the introduction of effective control measures, or closure of the industry/process), the beginning and end of the REP was adjusted to correspond to this. For example if point 5 occurred before UL, 5 replaced UL as the end of the REP. If 5 preceded LL, the REP disappeared. In this case the exposure failed inclusion test 2 (see Appendix 3).

The *source REP*, for the study chosen for the estimate of relative risk, was calculated as:

Source REP =
 [source risk period minus maximum cancer latency] to [source risk period minus minimum latency].

The source study risk period is the period of case collection for a case-control study or follow-up for a cohort study. The *target REP*, used to estimate AF was calculated as:

Target REP = [target year minus maximum latency] to [target year minus minimum latency].

The target year was the year for which an AF was to be estimated. This was the most recent year for which cancer mortality, or registration data, was available (currently 2004), or it can be set in the future, if predictions based on current exposures are needed.

For example, to estimate the burden of cancer in the year 2004 for a solid tumour for which latency of between 10 and 50 years has been assumed, the target REP would be 1955 to 1994. However, the best source study may have been carried out on cases collected around 1990 representing a source REP of 1941 to 1980. If the proportion of those exposed (if it is a population-based study), and the level of that exposure had remained constant over time this would not be a problem. If this was not the case adjustments to the estimates of the proportions exposed (overall and at the 'high' and 'low' exposure levels) were needed.

Where external estimates of proportions exposed were used, such as from numbers employed in an occupation or industry in GB, these were chosen to apply to the relevant exposure period or, in the case of the CAREX estimates, adjusted to apply to this period (see Appendix 6 below).

In practice, very little information was available on latency for individual cancer sites and cancer/exposure pairings. Therefore for the sake of clarity a single target relevant exposure period for the solid tumours, and a separate single target REP for the haematopoietic cancers were used throughout, unless the period was shortened due to specific changes to the individual exposures. For the target year of 2004 these were as follows. Data used in the calculation of AF specific to each standard REP were given in Table 1 in Section 3.5.

Solid tumours: The REP was defined as 1955 – 1994, based on a maximum latency of 50 years and a minimum of 10 years, and with a 'peak' at about 35 years (i.e. the highest number of solid tumours leading to current deaths or registrations related to exposures in the early 1970s).

Haematopoietic neoplasms: The REP was defined as 1985 – 2004, based on a maximum latency of 20 years and a negligible minimum, with a 'peak' at about 15 years (i.e. the highest number of solid tumours leading to current deaths or registrations relate to exposures in the early 1990s).

Appendix 6: Calculating the attributable fraction

A6.1 Population-based studies – Miettinen’s estimator

For population-based studies, as long as the ‘portability’ can be assumed, an estimate of the proportion of the population exposed, or of the proportion of cases exposed, was obtained directly from the study using the exposure distribution for the controls and the cases respectively. The cases had to represent a random sample of all the cases in the population. Using a population-based study has the advantage that Miettinen’s estimator of the attributable fraction can be used. This is unbiased when adjusted relative risk estimates are being used as is often the case in more recent studies (Benichou, 2001).

Miettinen’s formula for the attributable fraction (AF):

$$(6.1.1) \text{ AF} = \text{Pr}(E|D) * (\text{RR} - 1) / \text{RR}$$

where $\text{Pr}(E|D)$ = proportion of cases exposed

This formula requires an estimate of the proportion of the cases exposed, which was not available other than from the study data.

If the alternative Levin’s formula (see below) is used with adjusted RR estimates, the results will be biased (Greenland, 1984; Rockhill, Newman and Weinberg, 1998) but the direction and degree of this bias can be determined, and will not normally be significant in comparison with other potential sources of bias in the AF calculation.

AFs calculated using Miettinen’s formula can also be combined across ‘sufficient cause’ sets of exposures (see Appendix 7), where interaction and / or confounding between exposures is thought to be a problem. These sets also can be partitioned without introducing bias into the resulting AF (Bruzzi et al., 1985).

A6.2 Industry-based studies – Levin’s estimator, and estimating the proportion in the population exposed

For industry-based studies, the proportion of the population exposed was calculated from independent data. Here the study members were not considered representative of the population as a whole. Indeed, an industry cohort was in certain circumstances considered 100% exposed, depending on the definition of exposed versus unexposed. Two estimates were needed from independent data to arrive at the proportion of the population exposed. The first was the *proportion in the workplace exposed* to the substance of interest (i.e., above a defined baseline level, to match the definition of exposure used for the source study estimate of RR). This was in practice very difficult to obtain, as it required numbers exposed and unexposed in the workplace, as well as substance level measurements to define the exposure. Apart from the possibility of sporadic coverage in the NEDB, CAREX was the only known comprehensive source for this part of the estimate. CAREX gave numbers exposed in 1990-93, which were adjusted to the relevant exposure period. The adjustment equation for this was as follows:

$$(6.2.1) N_{e(\text{adj})} = N_e * \{N_{\Sigma\text{SIC}}(\text{mid year}) / N_{\Sigma\text{SIC}}(\text{ave1990-93})\}$$

where N_e = CAREX /NEDB numbers exposed [1990-93 applies to CAREX data]
 $N_{\Sigma\text{SIC}}(\text{mid year})$ = ABI/CoE or LFS numbers employed in the relevant SIC or other classification codes in the middle year of the REP.

In practice, for the sake of clarity, a limited number of adjustment factors for N_e were used throughout, to adjust for the higher numbers employed in manufacturing industry and mining and quarrying during the ‘standard’ relevant exposure period. This pre-dated the fall in employment in this sector, which occurred between the late 1970s and early 1990s. This adjustment factor was 1.4 for men and 1.5 for women. Other factors were used to account for the lower numbers employed in the service sector in the 1970s compared to the CAREX base-line years (see section A6.3 below).

The other estimate needed was the *numbers ever employed in the target relevant exposure period*. This was obtained from CAREX or national employment data (usually the Census of Employment and its successors, or the Labour Force Survey (LFS)), using where possible the relevant Standard Industrial Classification (SIC) or Standard Occupational Classification (SOC) codes, and calculated using a formula that took into account staff turnover during the REP. Staff turnover was estimated from the LFS ‘length of service of employees’ data (see section A6.4 below). The ‘turnover equation’ used to calculate numbers ‘ever employed’ in the target REP (using life table estimates of the proportions surviving to the target year 2004) was as follows:

$$(6.2.2) \quad N_{e(\text{REP})} = \sum_{i=a}^{i=b} l_{(\text{adj}15)_i} * n_0 / (R-15) + \sum_{j=c}^{j=d} \{ l_{(\text{adj}15)_j} * n_0 * \text{TO} \}$$

where

n_0 = numbers employed at the beginning of the REP [= $N_{e(\text{adj})}$ for the CAREX and NEDB data from equation (6.2.1)], for the relevant SIC or SOC codes (industry based studies). Estimates based on adjustment to the early 1970’s for the solid tumour REP are used as (1) no earlier data is readily available on which to base backdated adjustment factors, and (2) employment patterns are thought not to have changed significantly between the mid 1950s and early 1970s.

TO = staff turnover per year, from LFS estimates of duration of employment

R = retirement age (65 for men, 60 for women)

$l_{(\text{adj}15)_i}$ = the proportion of survivors to exact age i of those alive at age 15 who are assumed to be subject throughout their lives to the mortality rates experienced in the three year period to which the relevant Interim Life Table relates⁷, (1980-82 for the solid tumour REP, 1994-96 for the short latency REP). The proportions are sex specific.

a to b = age range achieved by the original cohort members by the target year (age 65 to 100 for the solid tumour REP; age 35 to 84 (M) or 79 (F) for the short latency REP)

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

Assumptions underlying this estimator were (1) there was an even distribution of ages across the cohort in its first year, and (2) that recruitment to the cohort through turnover was into the youngest age (15) only.

The assumption that recruitment into the cohort was only at age 15 is unlikely to be realistic. An alternative equation, that assumes that recruitment into the cohort is evenly spread across a range of ages, is given below. Assuming recruitment across all ages (15 to retirement age) results in greatly reduced estimates of surviving exposed numbers, and is also probably unrealistic.

⁷ Government Actuary’s Department, Interim Life Tables, at http://www.gad.gov.uk/Life_Tables/Interim_Life_Tables.htm. $l_{(\text{adj}15)_i}$ is calculated as $l_x(\text{age } i) / l_x(\text{age } 15)$

$$(6.2.2^*) \quad N_{e(\text{REP})} = \sum_{i=a}^{i=b} I_{(\text{adj}15)_i} * n_0 / (R-15) \} \\ + \sum_{k=0}^{k=(\text{age}(u)-\text{age}(l)+1)} \sum_{j=c+k}^{j=d+k} \{ I_{(\text{adj}15)_j} * n_0 * \text{TO} / (\text{age}(u)-\text{age}(l)+1) \}$$

where age(u) and age(l) = upper and lower recruitment age limits (e.g. 24 and 15)

In practice the values of a, b, c and d were chosen as indicated above for a REP of 1955 to 1994 for the solid tumours, and 1985 to 2005 for the haematopoietic cancers (see Appendix 5 above), and TO was set to a standard for each main employment sector. The derivation of TO is given in section A6.4 below.

The equation can be represented as a single factor acting as a multiplier for n_0 , calculated by setting n_0 to one in the above equation, so that the factor varies only with TO (see Section A6.4 below). Separate values for TO and therefore this multiplier were used for men and for women, as their employment turnover rates differed.

The numbers ever employed in the target REP from equation (6.2.2)) were divided by an estimate of *numbers ever of working age during the REP* (calculated from population estimates by age cohort in the target year⁸) to give the proportion of the population who were employed in that sector of the industry (or occupation) during the relevant exposure period.

Numbers ever of working age during the REP were determined as follows⁹:

$$(6.2.3) \quad N_{p(\text{REP})} = \sum_{\text{age}(j=a)}^{\text{age}(j=b)} (N_{\text{age}(j)}) \text{ for the target year (2004)}$$

where $N_{\text{age}(j)}$ = population in age group j in the target year
a = youngest working age group (15-24) + (TY-UL)¹⁰
b = oldest working age group (e.g. 60-64) + (TY-LL), or age group (90+) if this is reached first¹¹
LL = first year of REP
TY = target year (2004)
UL = last year of REP

Again, this value becomes a constant across cancer/exposure pairings when based on the 'standard' relevant exposure periods. For the solid tumour REP, the numbers were 19.2 million men and 20.9 million women. For the short latency cancers REP, the numbers were 20.9 million men and 19.8 million women.

⁸ Mid-2004 Population Estimates: Great Britain; estimated resident population by single year of age and sex; Table 2. <http://www.statistics.gov.uk/statbase/ssdataset.asp?vlnk=9083&More=Y>

⁹ This is the population of working age and over, up to the maximum age that someone retiring during the REP could have reached by the last year of the REP (1994 for solid tumours), moved forward in age from then to the target year (2004).

¹⁰ i.e. (25-34) for the solid tumour REP, (15-24) for the short latency REP

¹¹ i.e. (90+) for the solid tumour REP, (80-84) for the short latency REP for males, (75-79) for females

This proportion was then multiplied by the estimate of *proportion in the workplace exposed* to arrive at an estimate of the *proportion in the population exposed*:

$$(6.2.4) \Pr(E) = \Pr_{wk}(E) * N_{e(REP)} / N_{p(REP)}$$

where $\Pr(E)$ = proportion in the population exposed
 $\Pr_{wk}(E)$ = proportion in the workplace exposed

$\Pr_{wk}(E)$ was set to unity where CAREX or NEDB data of exposed numbers is used.

Where no estimate was available for the proportion in the workplace exposed, the proportion ‘ever employed’ in that industry sector (or occupation category) during the relevant exposure period had to suffice, but it was likely to have overestimated actual substance exposures. Again the key elements here were the match between the definition of the exposed versus the non-exposed used for the ‘best study’ RR estimate, and the nearest approximation that can be obtained from the independent SIC code based GB national employment estimates (including CAREX).

Levin’s formula was then used to calculate AF.

$$(6.2.5) AF = \Pr(E) * (RR - 1) / \{1 + \Pr(E) * (RR - 1)\}$$

where $\Pr(E)$ = proportion of the population exposed

All the data required for these calculations were entered onto a single spreadsheet for each cancer site.

A6.3 Adjusting CAREX exposed numbers to the relevant exposure period

There have been some significant changes in the structure of industry and therefore patterns of employment in GB in the last 35 years. Manufacturing industry and employment in mining and agriculture in particular have declined, and numbers employed in service industries have increased. As most cancers have a relatively long latency period, these changing patterns of employment needed to be taken into account when estimating the numbers exposed to occupational carcinogens.

Figures 6.3.1 and 6.3.2 illustrate the changes in employment over the time periods for which we had data, for grouped main industry sectors which were comparable across the different classifications used over the years, for the LFS and industry-based surveys. Due to changes in the way the ‘Broad Industry’ Groups were defined between SIC68 and SIC80, mining and quarrying and energy and water industries were grouped with manufacturing. Data were used for 1971 to 2004 based on the CoE and its successors, and for 1979 to 1997 based on the LFS, which includes estimates for the self employed.

Both data sources indicated a sharp decline in employment in manufacturing industry and mining between the late 1970s and early 1990s. The LFS included the self-employed, and was therefore a better source for overall employment and for the construction industry. The LFS data for agriculture also looked more realistic when matched to numbers for farmers and their employees from the June Agricultural Census.

Figure 6.3.1

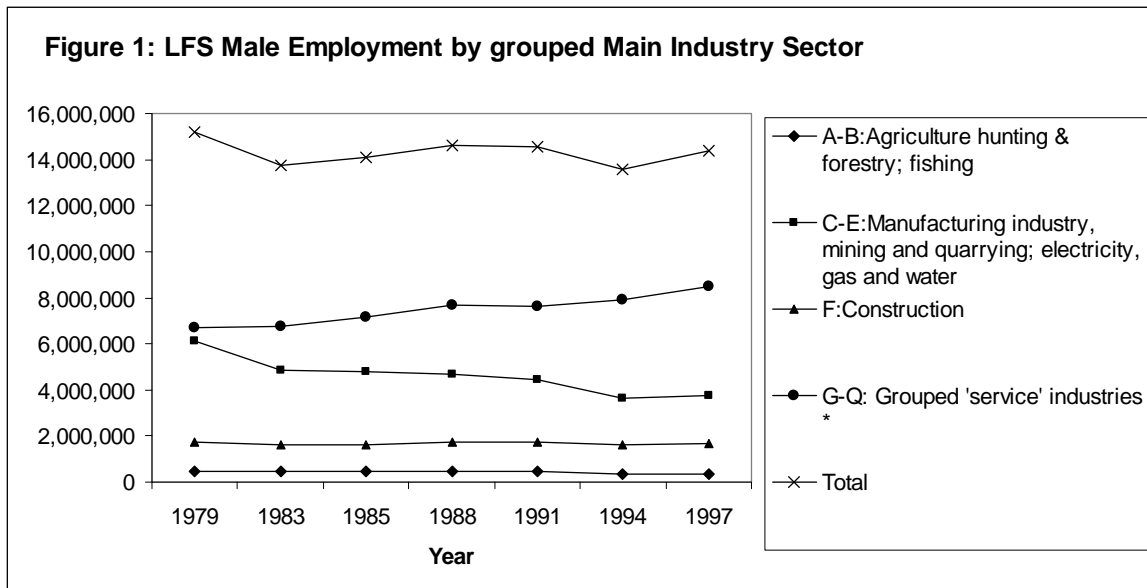
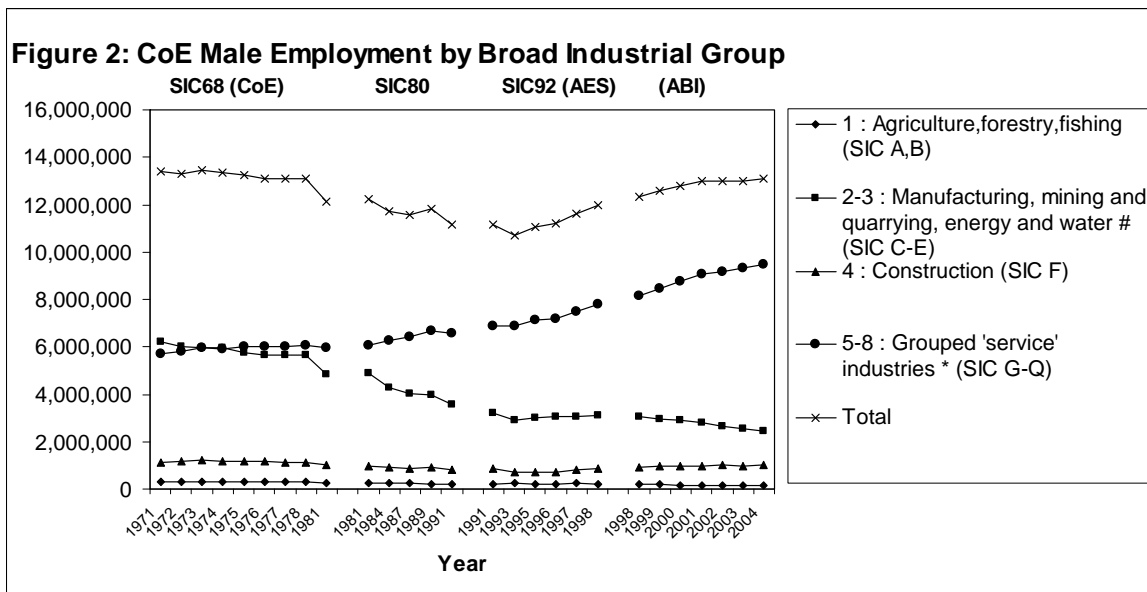


Figure 6.3.2



SIC68: 02.Mining and quarrying, 2.Gas, electricity and water, 3.Manufacturing industries.
 SIC80: 2.Energy and water supply, 3.Manufacturing industries.
 * SIC68: 5.Distributive trades, 6.Transport/communication, banking, finance, 7.Public administration and defence, 8.Miscellaneous services.
 SIC80: 5.Distribution, hotels/catering, repairs, 6.Transport/communication, banking, finance, 7.Public administration and defence, 8.Other service industries

The CAREX data were for 1990-93.

The ‘standard’ relevant exposure period for solid tumour cancers was 1955 to 1994, with current cancers taken as relating ‘on average’ to exposures in the middle of this period in the early 1970s. We therefore needed to adjust the CAREX data to take account of these changing employment levels between the early 1990s and early 1970s. For the shorter latency neoplasms the ‘standard’ relevant exposure period was 1985 to 2004, peaking in the early 1990s and therefore requiring no adjustment.

To decide on the factors to use to backdate CAREX data for the solid tumour cancers, ratios of numbers employed for the broadly grouped industry categories that were consistent across the classifications were used. These ratios along with the data supporting the calculations are in Table A1 in Annex 1. Although not available back to 1973 (the mid-REP year chosen for the adjustment), the LFS was the better source of data to use for factors applying to the main industry groups A-B (agriculture, forestry and fishing), F (construction) and for the total. All these factors were close enough to one (implying no significant change in numbers between the early 1970s and early 1990s) and were therefore ignored, at least for men. However the increase in paid employment amongst women has resulted in rising employment in all these sectors. CAREX numbers were not separately available for women, but where assumptions were made for women’s exposure based on the CAREX data, the adjustment factors in Table 6.3.1 were applied to deflate the numbers of women who would have been exposed in the early 1970s. For the grouped service industries, Figures 6.3.1 and 6.3.2 indicate rising numbers in these sectors across the period. The CoE adjustment factor for 1973 to 1991 was nearly equal to the LFS factor for 1979 to 1991, and so this factor was used. For women the rise in employment was again significant; and the LFS factor has been used for the sake of consistency. The most important factor was for the combined manufacturing, mining and quarrying and energy and water industry sectors, where there has been a substantial fall in the numbers employed, most of which has occurred since 1978 (see Figure 6.3.2); this indicated that the LFS factor was the preferred choice as it took account of the self-employed. The factor was 1.4 for men, and 1.5 for women.

The factors used are given in Table 6.3.1:

Table 6.3.1: Factors used for adjusting CAREX exposed numbers to the relevant exposure period (solid tumour cancers)

	Main Industry Sector		CAREX adjustment factor
Men	A,B	Agriculture, hunting and forestry; fishing	1
	C,D,E	Mining and quarrying, electricity, gas and water; manufacturing industry	1.4
	F	Construction	1
	G-Q	Service industries	0.9
		Total	1
Women	A,B	Agriculture, hunting and forestry; fishing	0.75
	C,D,E	Mining and quarrying, electricity, gas and water; manufacturing industry	1.5
	F	Construction	0.67
	G-Q	Service industries	0.8
		Total	0.9

Where the CAREX exposure counts covered more than one group of main industry sectors, the CAREX numbers were partitioned between the grouped sectors and adjustment factors were applied separately.

A6.4 Employment Turnover Calculation (TO)

The calculation used LFS data for the ‘length of time with current employer’, with eight ‘duration of employment’ categories from <3 months to >20 years. The data used was available by occupation codes and by SIC main industry sector four digit codes and by year for 1983, 1984, 1991, 1998, 2000 and 2003.

The year chosen for the calculation of TO needed to be as close as possible to the mid-point of the relevant exposure period for the cancer. For the ‘standard’ relevant exposure period for solid tumours (1955 to 1994), 1984 was chosen as the earliest available year with SIC coding consistent through to 2003. As the LFS is a sample survey, estimates based on those codes which represented small numbers of workers would have been unstable. Consequently, an estimate for broad industry sectors has the advantage both of being stable and providing consistency across multiple exposures. The grouped main industry sectors used for the CAREX adjustment factors were also used for the turnover estimates, so that the partitioning of the numbers exposed was the same for both factors.

As the calculation was based on ‘current employer’, an estimate had to be made of how long an employment was expected to last, and from this calculation the numbers were re-allocated to longer service categories. Numbers in each category were re-distributed to the other categories based on the distribution of numbers in all the other categories, averaged across the relevant future years, in this case 1984, 1991 and 1998.

To do this for category [k] of the eight categories, the n_k employees were re-distributed across all categories above and including k, according to the proportions of employees in these categories from k upwards for the average of the years 1984-1998.

$$(6.4.1) \quad n_{k,adj} = \sum_{j=k}^8 (n_j * p_j / (\sum_{i=k}^8 p_i - \sum_{i=k}^{j-1} p_i))$$

where $n_{k,adj}$ = adjusted count in category k
 n_j = numbers in length of employment category j
 p_j = numbers in category j averaged over 1984, 1991 and 1998

Annual turnover was then calculated as the weighted sum of the inverse of the length of employment category mid-points (m_j), the weights were the adjusted employee counts as a proportion of the total employees:

$$(6.4.2) \quad TO = \sum_{j=1}^8 (1/m_j * n_{j,adj} / \sum_{i=1}^8 n_i)$$

where TO = employment turnover per year
 m_j = mid-point of category j

To calculate TO for only employment categories of 1 year and over, or of 5 years and over, the categories below these ‘cut-offs’ were excluded in the summation for TO in equation (6.4.2) above.

The turnover factor was calculated using a modification of the Global Burden of Disease formula in Nelson et al., (2005)¹².

$$^{12} \text{ Turnover factor} = N_{e(REP)}/n_0 = \sum_{i=a}^{i=b} \{I_{(adj)15}{}_i / (R-15)\} \\ + \sum_{k=0}^{k=(age(u)-age(l)+1)} \sum_{j=c+k}^{j=d+k} \{I_{(adj)15}{}_j * TO / (age(u)-age(l)+1)\}$$

The estimates for TO used in the current study are listed in Table 6.4.1 below, in bold. The estimates for those employed for at least one year were used as they excluded those possibly exposed for very short periods that would not normally be considered at risk. These estimates also best matched other estimates derived by other researchers (see below). Other estimates are given for comparison only.

Table 6.4.1: Estimates for employment turnover (TO)

Sector		All length of employment categories		Employed \geq 1 year		Employed \geq 5 years	
		Annual turnover	Turnover factor over 40 years	Annual turnover	Turnover factor over 40 years	Annual turnover	Turnover factor over 40 years
A,B : Agriculture, hunting and forestry; fishing	M	0.14	5	0.09	3	0.05	2.1
	F	0.19	8	0.10	4	0.05	2.4
C-E: Mining and quarrying; manufacturing; electricity, gas and water	M	0.12	5	0.09	4	0.06	2.3
	F	0.22	8	0.14	6	0.06	2.7
F: Construction	M	0.22	8	0.13	5	0.06	2.3
	F	0.25	10	0.16	6	0.06	2.5
G-Q: Grouped 'service' industries	M	0.16	6	0.11	4	0.06	2.3
	F	0.24	9	0.15	6	0.06	2.7
Total	M	0.15	5	0.10	4	0.06	2.3
	F	0.23	9	0.15	6	0.06	2.7

The turnover calculation based on the estimated duration of employment overestimates numbers exposed, as the duration represents employment in one job, rather than the industry as a whole.

where $N_{e(REP)}$ = numbers ever exposed during the relevant exposure period (REP)
 n_0 = numbers employed at the beginning of the REP
TO = staff turnover per year
R = retirement age
 $l_{(adj)15i}$ = the proportion of survivors to exact age i of those alive at age 15 who are assumed to be subject throughout their lives to the mortality rates experienced in the three year period 1980-82 for the solid tumour REP, 1994-96 for the short latency REP.
a to b = age range achieved by the original cohort members by the target year
c to d = age range achieved by the turnover recruited cohort members by the target year
age(u) and age(l) = upper and lower recruitment age limits (e.g. 24 and 15)

This may be partially corrected as follows, by using the LFS counts of ‘has respondent’s occupation changed over the last year?’ Each count in the first three length of employment categories (i.e. employed in current job for less than one year) is multiplied by the ‘yes’s’ to this question as a proportion of the sum of total employees in these three categories, and the result is used to replace the length of employment category counts n_j in equation (6.4.1). To illustrate the effect of this, I have assumed in the calculations below that 10% of employed males and females answered yes. The values for TO and its associated turnover factor, under the same headings as above, are then as follows:

For 10% answering yes to ‘has respondent’s occupation changed over the last year’:

Total	M	0.11	4	0.10	4	0.06	2.4
	F	0.16	6	0.14	6	0.07	2.8

For 5% answering yes the results are:

Total	M	0.10	4	0.10	4	0.06	2.4
	F	0.14	6	0.14	5	0.07	2.8

These results are very close to the estimates for those employed for at least one year, indicating that this additional refinement does not add to the reliability of the AF calculations.

Other authors’ estimates:

HSE (2002) estimated a 10-fold lifetime (over 40 years) turnover factor (implying a 25% annual rate), for demolition and asbestos-removal workers (consistent with data on individuals having statutory asbestos medicals as asbestos removal workers over the previous 14 years). 5-fold (12.5%) and 2.5-fold (6%) turnover factors were determined for maintenance and other building work respectively. These were based on the LFS, which provides estimates of time with current employer, and also on whether the respondents occupation had changed over the last year.

A factor of four has been used by NIOSH on a cross-sectional exposure survey from early 1981, to estimate the number of workers ever exposed to lung carcinogens as of 1997, in order to estimate AFs in 1997 (communication from Kyle Steenland¹³). The factor of four came from assuming a 10% turnover per year, along with a 40 year period of interest (1957-1997); a given age distribution at baseline in 1957; an average duration of exposure of 10 years (log normal distribution); and a mortality loss of 20% over the 40 year period. This scenario assumed an average latency of approximately 20 years. This estimate has been used in the Global Burden of Disease methodology (Nelson et al., 2005).

Dreyer et al., (1997) used a turnover assumption that 85% of blue collar production workers were engaged in the same trade for at least one year during the 15 year period 1970-84, implying a yearly turnover of the workforce of 15%. They also assumed 1970-84 was the REP for the cancer pattern in the year 2000.

A6.5 Partitioning Pr(E) between ‘high’ and ‘low’ exposed groups

A partition by level of exposure may need to be in two dimensions, according to industry/occupation on the one hand and time period on the other.

¹³ GB Burden of Occupational Cancer: Summary Report of Workshop Held on the 22nd and 23rd November 2004 in Manchester: HE/05/03

A6.5.1 Partition by industry/occupation

CAREX counts are industry coded and the CAREX numbers for the ‘higher’ exposed group were grouped according to which industries potentially had ‘average’ exposures similar to those in the study(ies) from which the ‘higher’ risk estimates were obtained (i.e., based on these codes). The rest were allocated to the ‘lower exposed’ group, unless there are codes considered not to belong here with exposure scenarios also not matching the RR estimates that were obtained for the ‘higher’ group. Additional ‘best study’ RRs were then needed for these exposure scenarios, and they were treated as an additional group. In some cases CAREX numbers exposed were allocated to a ‘background’ exposed group, and assumed to have a negligible risk, RR=1.

The ‘higher’ and ‘lower’ counts were also split between the grouped main industry sectors, so that separate CAREX adjustment and turnover factors were applied (see Appendices A6.3 and A6.4) before Pr(E) was calculated for the group as a whole.

A6.5.2 CAREX partition by sex

CAREX data are not available separately for men and women. However they do not include white collar workers in the industry groups covered. To split the CAREX numbers between men and women, if there are thought to be women exposed, the male to female proportions were obtained as follows.

- 1) If the risk estimate studies include very few women and it is thought that this also applies to the UK then the CAREX figures were assumed to be all men.
- 2) If the risk estimate studies include more than just a few women then the same proportion as that in the studies was assumed for CAREX if the UK population was thought to be similar to the study population.
- 3) Alternatively the AES or LFS was used to get the ratio of males to females for all those employed in the exposure scenario and then, using knowledge of the jobs that women do in the industry, a judgement was made about what proportion of the women are exposed to levels of concern i.e. not just background, and the CAREX figures were apportioned accordingly. For example if the ratio of men to women is 50:50 overall in an industry group from the AES/LFS but it is thought that 50% of the women will have had little or no exposure (and were therefore excluded from CAREX), and only 50% will have any exposures of concern then the ratio of M:F in the CAREX data is 75:25.
- 4) In practice an occupation by industry cross tabulation by sex of numbers over age 16 employed or self-employed in Great Britain has been used, from the 1991 Census. From these data proportions were estimated, using mainly standardised sets of the major occupation groups, for the broad industrial groups that were used for the CAREX adjustment and turnover factors described above. The estimated male percentages are given in Table A2 in Annex 1. Separate percentages are available for all workers in the broad industrial groups and for appropriate combinations of these groups. The most useful are for “blue collar” workers, defined as those in Standard Occupational Classification major groups 5, 8 and 9 which cover skilled trades, shop floor and transport operatives; for “white collar” workers, SOC major groups 1, 2 and 4, managerial, professional, administrative and secretarial; and for 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 is used, with flexibility to use percentages for single rather than grouped SOC major groups if exposed numbers are clearly concentrated in these occupations. Transport operatives (SOC group 8) in the service industries (industry sectors G-Q) were an example of the appropriate use of a single SOC group for exposure to diesel engine exhaust.

A6.5.3 Partition by time period

If it was known that the exposure level had changed sufficiently during the target relevant exposure period to move from ‘high’ to ‘low’ exposure, the REP needed to be split between these periods to calculate Pr(E). This was done by partitioning the estimate for the numbers ‘ever employed’ in the REP in the industry/occupation-defined group ($N_{e(REP)}$) between the ‘high’ and ‘low’ exposure periods ($N_{e(REP)1}$ and $N_{e(REP)2}$) according to the age ranges achieved by the target year by the members of the cohorts in each successive period. Note that members of the original cohort were only counted once, in the first period. Examples are given below for the solid tumour REP (1955-1994) split between the periods 1955-62 and 1963-94.

$$\text{From (6.2.2)} \quad N_{e(REP)} = \sum_{i=a}^{i=b} l_{(adj15)i} * n_0 / (R-15) \} \\ + \sum_{k=0}^{k=(age(u)-age(1)+1)} \sum_{j=c+k}^{j=d+k} \{ l_{(adj15)j} * n_0 * TO / (age(u)-age(1)+1) \}$$

Where

a to b = age range achieved by the original cohort members by the target year (age 65 to 100 for the solid tumour REP; age 35 to 84 (M) or 79 (F) for the short latency REP)

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

1955-62

$$(6.5.3.1) \quad N_{e(REP)1} = \sum_{i=a}^{i=b} l_{(adj15)i} * n_0 / (R-15) \} \\ + \sum_{k=0}^{k=(age(u)-age(1)+1)} \sum_{j=c1+k}^{j=d1+k} \{ l_{(adj15)j} * n_0 * TO / (age(u)-age(1)+1) \}$$

Where

a to b = age 65 to 100

c1 to d1 = age 57 to 64

1963-94

$$(6.5.3.2) \quad N_{e(REP)2} = \sum_{k=0}^{k=(age(u)-age(1)+1)} \sum_{j=c2+k}^{j=d2+k} \{ l_{(adj15)j} * n_0 * TO / (age(u)-age(1)+1) \}$$

where c2 to d2 = age 25 to 56

Note that the denominator (i.e., the numbers ever of working age during the REP) remains the same as calculated for the whole period (1955-1994 for the solid tumours) and was not split between the two periods. The denominator also remained the same if an exposure was deemed to have ceased before the end of the REP for the cancer, although the numerator (i.e., the numbers ‘ever exposed’) was based on a shorter time period (e.g. 1955-1962).

Appendix 7: Combining AF across exposures

To sum the exposure levels for a single exposure (h, which was usually ‘high’ and ‘low’ for this study) they must have a common reference level. The equation used for Levin’s estimator was:

$$(7.1) \quad AF = \{\sum_h \Pr(E_h)(RR_h - 1)\} / \{1 + \sum_h \Pr(E_h)(RR_h - 1)\}$$

and for Miettinen’s equation was:

$$(7.2) \quad AF = \sum_h \Pr(E_h|D) (RR_h - 1) / RR_h = \Pr(E|D) - \sum_h \Pr(E_h|D) / RR_h$$

Only exposures that did not overlap in the workforce, as identified in the cancer ‘exposure map’, or whose disease incidence rates were known to be additive, were summed directly. It is also possible to sum ‘sufficient cause’ (Hoffmann et al., 2006) sets of exposures. A ‘sufficient cause’ set relevant to this study is the set of exposures that may overlap and not be independent of one another, but when taken as a set do not overlap or interact with other exposures or sets of exposures. However, this requires a single estimate of RR and Pr(E) or Pr(E|D) for the whole set of exposures. This may be available from a population-based study, but it will be very difficult in practice to separate out the sufficient cause classes and therefore the contributions of the individual exposures within the set. As HSE needed to know the relative contributions of the different occupational exposures to prioritise interventions, separate AFs were required for each cancer/exposure pairing. These may have to be obtained from an alternative ‘best study’ to the one chosen for the ‘sufficient cause’ set, so leading to inconsistency between the individual and combined AFs.

The approach usually adopted was therefore to partition sections of the workforce known to be subject to multiple exposures between each exposure, allocating occupation/industry codes to the ‘highest risk’ or dominant exposures. The AFs were then directly summed. There were various strategies to achieve this and examples of their use are applied to the analysis for lung cancer.

1. Exposure scenarios were excluded entirely, if they wholly overlapped with another dominant exposure. Exposures shaded grey in Figure 2 (Section 3.2), were in general excluded for this reason, and did not have an AF calculated separately.
2. Alternatively, where the exposed populations only partially overlapped, exposed numbers were partitioned between the overlapping exposures. Then the AFs for the exposures concerned were summed directly. This strategy was adopted for the exposures circled in hashed lines in Figure 2 in Section 3.2 for lung cancer. Although these clearly overlapped in the working population with other exposures (as indicated by the joining lines in this figure), AFs were calculated for a restricted group of workers only, which (as far as possible) did not include workers in the populations for other exposures with which they overlapped. Tin miners, steel foundry workers and mineral oils (printers only) were treated in this way. The exposed populations (‘iron and steel basic industries’ as defined in the CAREX data, ‘printers’ as narrowly defined by occupation from the LFS, and tin miners from an established UK cohort) were excluded from the other exposures with which they overlapped.
3. If it was known that a set of overlapping exposures were independent and their joint effect on initiating or promoting cancer was multiplicative, i.e. $RR_{(\text{exp 1 and exp 2})} = RR_{(\text{exp 1})} * RR_{(\text{exp 2})}$, the exposed numbers were not adjusted. In this case the AF 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:

$$(7.3) \quad AF_{\text{overall}} = 1 - \prod_k (1 - AF_k) \text{ for the } k \text{ exposures in the set.}$$

The exposures circled in bold in Figure 2 (Section 3.2) were treated in this way (termed 'AF product' sets). The AF for the set as a whole was then summed with other non-mutually overlapping 'AF product' sets or individual exposure AFs.

For lung cancer, asbestos, silica, diesel engine exhaust, PAHs and occupation as a painter overlapped as potential cancer causing exposures in the construction, quarrying and potteries industries and amongst drivers and those employed in personal and household services (amongst other possible exposure scenarios). These exposures were therefore treated as a set, with the component exposures assumed to act independently of one another. Likewise arsenic, nickel, chromium, lead and cadmium were all found in the smelting, refining, alloy, plating and battery industries, and were also assumed to act independently of one another. These exposures were not found in the industries listed for the asbestos-silica-diesel-PAHs-painters set, except lead exposure, in plumbers, which overlapped with asbestos exposure in construction, and in printers with mineral oils. Testing the results in empirical terms was useful in these circumstances. Transferring exposure to lead to either the asbestos set, or into a product set with mineral oils, did not, however, affect the overall AF result by more than 0.1% in absolute terms. Exposure to silica and to arsenic also overlapped in the glass industry.

In the case of lung cancer, welders were a difficult case, as this exposure overlapped with two 'AF product' sets, and it was not possible to identify and exclude 'welders' specifically from the CAREX numbers for its overlapping exposures. In practice it made a difference of only 0.1% in absolute terms to the overall AF if the AF for welders was summed directly, or treated as a member of either of the 'AF product' sets. No attempt was made to exclude welders from other exposed number estimates.

There were two alternative ways to combine the two 'AF product sets' and the individual exposures with which they overlapped (welders and mineral oils in printing). If it was thought that substantial proportions of the exposed workers were involved (which may be the case for welders, and for exposure to lead in the construction industry) the 'complement of the product of complements' equation was used; otherwise a direct sum was probably more appropriate (note that both are compromises). The results were very similar in practice.

ETS and indoor radon exposure at work also overlapped in high radon areas, and so were treated as a third product set. No overlap was assumed with other exposures and product sets.

4. The exposures were also summed directly if the exposure disease rates were known to be additive.

5. The AFs for exposures with no overlaps with others (no line connections in the exposure map) were summed directly. The exposures with no remaining overlaps in Figure 2 (e.g., dioxin and beryllium for lung cancer) were therefore summed directly with one another, and with the 'AF product' sets.

Appendix 8: Random error confidence intervals

For each AF, a confidence interval that takes into account random error only was calculated. The equations for the various calculation methods and data sources are given below and were based upon methods published by Steenland and Armstrong (2006). However, it was acknowledged that the effect of bias needed to be accounted for. A full sensitivity analysis and description of the methodology for estimating confidence limits that incorporate bias, known as ‘credibility limits’, and examples of these will be the subject of a forthcoming paper.

Confidence interval for random error only

Methods used to derive a confidence interval for the AF that take account only of this type of error are well established. The simplest method is to use the upper and lower confidence interval bounds published with the relative risk estimate in the AF calculation to obtain a ‘naïve’ confidence interval for the AF (Greenland, 2004). This assumes no random variability in the estimate of the proportion of the population exposed [Pr(E)], which may be acceptable where this estimate is based on national census data. This was the method used for the majority of the current AF CI calculations.

Where the proportion of cases exposed [Pr(E|D)] was derived from the same case-control study source as the estimate for RR, and AF was calculated using Miettinen’s equation, a formula to estimate the variance and confidence limits is given by Steenland and Armstrong (2006), and Greenland (1987). The equation is:

$$(8.1) \quad \text{var}(\ln(1-\text{AF})) = \frac{(\text{AF})^2 / (1-\text{AF})^2}{\{n_{\text{nonexp-cases}} / n_{\text{exp-cases}} n_{\text{cases}}\}} \left[\frac{(\text{var} \ln(\text{RR}))}{(\text{RR}-1)^2} + \left\{ \frac{2}{n_{\text{exp-cases}}} (\text{RR}-1) \right\} \right]$$

Formula (8.1) is used to estimate the upper and lower confidence limits of the AF (ULAF and LLAF); these are $1 - \exp[\ln(1-\text{AF}) \pm 1.96(\sqrt{\text{var}(\ln(1-\text{AF}))})]$. This formula is used whether or not the RR was adjusted for confounding, assuming approximate homogeneity of the RR across the confounder strata. It is based on a single summary RR, and is used with a Mantel-Haenszel RR or an RR which was derived from a regression model. However, the exposure status of the cases in the source population has to be known. Formula (8.1) is used for data from case-control, cohort, or prevalence studies (Steenland and Armstrong; 2006).

Where the estimate of Pr(E) for the AF was obtained from an independent (survey) source and Levin’s formula was used the following equation, from Steenland and Armstrong (2006) and Greenland (2004) was used:

$$(8.2) \quad \text{St.dev.}(\ln(1-\text{AF})) = \frac{[(O * \text{RR} * \text{st.dev.} \text{RR})^2 + (O/T) \{ (\text{RR}-1)^2 + (\text{RR} * \text{st.dev.}(\text{RR}))^2 \}]^{1/2}}{*[(1-\text{AF})/(1+O)]}$$

where T = number in survey
O = Pr(E) / (1-Pr(E)).

This incorporates the variance of Pr(E) taken from an ancillary survey in the target population, and assumes that the RR in the source population is portable to the target population, and that there is no confounding in the target population. This formula can be used to obtain 95% Wald limits (which must be converted back to the original AF scale). It is unwieldy and becomes more complicated if either RR or O is estimated via a regression analysis or a complex survey (Steenland and Armstrong, 2006). In the case of CAREX estimates of the numbers exposed, the use of this formula was not appropriate, as the CAREX

estimates were not known to be based on a 'survey' of a target population for which T was known. Therefore 'naïve' limits were used, based on the random variability of RR alone when Levin's formula and CAREX exposed numbers were used to estimate AF. Where LFS data were used, the sample survey size of about 55,000 households resulted in a near zero second term in equation 8.2.

Even for complex study designs from which the RRs may be drawn, methods for calculating a confidence interval for the AF based on random error alone are generally tractable (Benichou and Gail; 1990). However, it is not straightforward to calculate a standard error (and therefore CI) for a combined AF based on more than one RR estimate.

Our preferred method in future will be to use 'credibility limits' based on Monte Carlo methods that allow for estimates of bias as well as random error. The methodology for these will be covered elsewhere, along with examples provided of their application.

Appendix 9: Alternative approach for asbestos exposure, based on the incidence of mesothelioma.

An alternative method of calculating an attributable fraction for lung cancer due to asbestos exposure was to estimate the number of excess deaths for lung cancer and for mesothelioma in job categories where there was known to have been exposure to asbestos. This was based on the incidence of mesothelioma amongst workers in these job categories. The ratio of the excess lung cancers to mesotheliomas was then taken as an indication of the numbers of lung cancers that were attributable to the exposure. The AF was then derived from this. The method was applicable only to asbestos exposure, as mesothelioma is considered uniquely to be caused by exposure to asbestos. This method had the advantage that it took account of the current impact that past levels of exposure to asbestos are known to be having on the incidence of cancer. There is a direct link to mesothelioma deaths, which are climbing rapidly, whereas lung cancer in general is declining due to the fall in the number of smokers. The link does, however, depend on the assumption that lung cancer has a similar pattern of latency to mesothelioma. Numbers of mesotheliomas in GB were obtained from the British Mesothelioma Register.

In populations with heavy asbestos exposures there have typically been at least as many (sometimes up to ten times as many) excess lung cancers as there have been mesotheliomas. The ratio depends on a range of factors - the most important of which are the types of asbestos (the highest ratios have been linked to chrysotile exposure), the level of exposure, age at exposure and smoking. Two lines of argument suggest that the ratio of asbestos related lung cancers to mesotheliomas in the British population as a whole is towards the lower end of the range of 1-10 estimated from the epidemiological studies. Firstly a study of lung cancer mortality in relation to indices of asbestos exposure and smoking habits in the west of Scotland suggested a ratio of around two asbestos lung cancers per mesothelioma for a region known to be associated with fairly high asbestos exposures. Secondly, analyses of mesothelioma deaths in GB by occupation and geographical area has suggested that substantial numbers of deaths may have arisen in workers other than those that were most heavily exposed¹⁴.

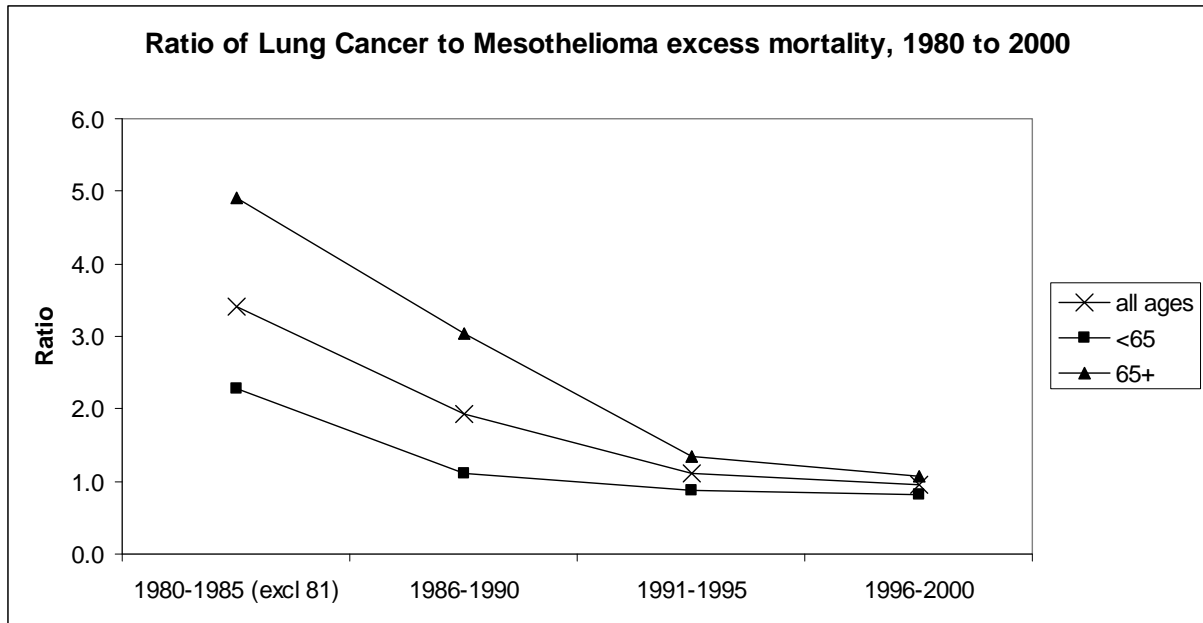
More recent evidence has suggested that this ratio may be at the bottom end of the range of 1-2. Asbestos is a more potent cause of mesothelioma than lung cancer. Smoking is also thought to interact with asbestos exposure in the causation of lung cancer. Thus going forward in time the ratio of lung cancers to mesotheliomas is likely to fall, because the mesotheliomas will increasingly be generated by low exposure levels of asbestos that are less likely to cause lung cancer and because smoking levels have fallen since the 1960s (factors that, together, mean fewer lung cancers per mesothelioma). A recent analysis of lung cancer mortality for the whole of GB (by occupational group in relation to indices of asbestos exposure and smoking habits) suggested that the ratio of asbestos related lung cancer to mesothelioma deaths was between two-thirds and one (Darnton et al., 2006). The current estimate of the ratio of mesothelioma to lung cancer deaths in jobs exposed to asbestos used here was 1:1, although this ratio, estimated for the period 1980-2000, may be decreasing as smoking and asbestos exposure levels both fall.

Occupations with exposure to asbestos were identified from their excess mortality from mesothelioma and were estimated using the proportional mortality ratio (PMR) data obtained from an analysis of death certificate data held on the HSE's Mesothelioma Register. Observed minus expected deaths (estimated as the average across all occupations) for mesothelioma and for lung cancer were calculated for each occupation, and summed across occupations that had a mesothelioma PMR with a 95% confidence interval excluding 100, to give an estimate of asbestos-attributable 'excess' deaths, for a particular time period (1980-2000, in five-year blocks, for this analysis). A suitable ratio of 'excess' lung cancer to

¹⁴ <http://www.hse.gov.uk/statistics/causdis/lungcan.htm>

mesothelioma deaths from this analysis was then applied to the count of mesothelioma deaths in the target year, to estimate attributable numbers. A ratio of 1:1 was chosen, as the ratio had clearly declined over the twenty year period as mesothelioma numbers have risen, although it appears to have levelled off at this ratio from 1990 onwards (see Figure 9.1 below). The AF was derived as the attributable numbers of lung cancers divided by total numbers of lung cancers in the target year.

Figure 9.1



7 ANNEXES

Annex 1

Table A1: Calculation of CAREX adjustment factors

Broad Industrial Group	CoE Employee counts (excludes self-employed)			For SIC68 to SIC80 conversion (on 1981 overlap data)	CAREX adjustment factor 1991 to 1973	CAREX adjustment factor 1991 to 1978 (for comparison with LFS adjustment)	LFS Employees + self-employed **		CAREX adjustment factor 1991 to 1979	LFS main industry group
	SIC68	1978	SIC80				1979	1991		
Male	1973	1978	1991				1979	1991		
1 : Agriculture,forestry,fishing	305,062	281,059	211,398	0.92	1.57	1.45	437,013	457,595	0.96	A-B:Agriculture hunting & forestry, fishing
2-3 : Manufacturing, mining and quarrying, energy and water #	5,984,066	5,633,601	3,554,973	0.99	1.71	1.61	6,096,181	4,468,216	1.36	C-E:Mining and quarrying, electricity, gas and water, manufacturing industry
4 : Construction	1,243,966	1,114,207	833,213	1.03	1.45	1.30	1,705,582	1,725,143	0.99	F:Construction
5-8 : Grouped 'service' industries *	5,942,661	6,046,512	6,570,033	0.99	0.92	0.93	6,690,244	7,600,644	0.88	G-Q: Grouped 'service' industries
Total	13,476,203	13,076,046	11,169,618	0.99	1.22	1.18	15,165,159	14,565,591	1.04	Total
Female										
1 : Agriculture,forestry,fishing	113,889	91,355	78,678	0.93	1.55	1.25	93,080	124,162	0.75	A-B:Agriculture hunting & forestry, fishing
2-3 : Manufacturing, mining and quarrying, energy and water #	2,375,996	2,169,121	1,443,922	0.98	1.68	1.53	2,392,842	1,625,075	1.47	C-E:Mining and quarrying, electricity, gas and water, manufacturing industry

4 : Construction	93,901	106,703	137,377	1.03	0.67	0.76	129,657	193,275	0.67	F:Construction
5-8 : Grouped 'service' industries *	6,119,483	6,782,548	8,739,414	0.99	0.71	0.78	6,943,567	8,549,368	0.81	G-Q: Grouped 'service' industries
Total	8,703,718	9,150,297	10,399,390	0.99	0.85	0.89	9,747,605	11,453,404	0.85	Total

SIC68: 02. Mining and quarrying, 2.Gas, electricity and water, 3.Manufacturing industries.
SIC80: 2.Energy and water supply, 3.Manufacturing industries

* SIC68: 5.Distributive trades, 6.Transport/communication, banking, finance, 7.Public administration and defence, 8.Miscellaneous services
SIC80: 5.Distribution, hotels/catering, repairs, 6.Transport/communication, banking, finance, 7.Public administration and defence, 8.Other service industries

** includes government schemes and family workers

Table A2: Numbers and percent of men and women in employment by industry and occupation

<i>Residents aged 16 and over - employees and self-employed, 1991 Census 10% sample</i>												
<i>Male % is given for single SOC major group only if it represents >20% of any one industry class in broad industry group</i>												
SIC Main Industry Sectors:	Agriculture, hunting and forestry; fishing			Mining and quarrying, electricity, gas and water; manufacturing industry			Construction			Service industries		
	A-B			C-E			F			G-Q		
	Male	Female	Male %	Male	Female	Male %	Male	Female	Male %	Male	Female	Male %
Standard Occupational Classification Major Groups:												
1 Managers and administrators	17,628	2,962	86%	55,941	14,129	80%	14,711	2,396		163,799	99,599	62%
2 Professional occupations	163	44		24,721	2,099		4,618	164		94,340	76,298	55%
3 Associate professional and technical occupations	77	65		22,497	7,154		4,558	398		75,062	93,838	44%
4 Clerical and secretarial occupations	125	1,195		20,182	42,506	32%	1,801	9,818		65,259	235,196	22%
5 Craft and related occupations	5,793	1,114		124,302	28,768	81%	96,292	745	99%	75,548	5,907	93%
6 Personal and protective service occupations	78	170		2,576	1,785		314	90		75,410	131,163	37%
7 Sales occupations	107	176		10,979	5,759		1,262	730		46,927	101,490	32%
8 Plant and machine operatives	708	515		95,967	41,146	70%	15,649	284		75,067	10,680	88%
9 Other occupations	11,052	3,243	77%	17,681	7,172	71%	17,331	873		52,174	90,917	36%
Occupation not stated or inadequately described	205	220		1,635	574		944	82		3,801	2,684	
TOTAL PERSONS	35,936	9,704	79%	376,481	151,092	71%	157,480	15,580	91%	727,387	847,772	46%
"White collar" Managerial, professional, administrative, secretarial: SOC groups 1,2,4	17,916	4,201	81%	100,844	58,734	63%	21,130	12,378	63%	323,398	411,093	44%
Associate professional, technical, personal and customer service occupations: SOC groups 3,6,7	262	411	39%	36,052	14,698	71%	6,134	1,218	83%	197,399	326,491	38%
"Blue collar" Skilled trades, shop floor and transport operatives: SOC groups 5,8,9	17,553	4,872	78%	237,950	77,086	76%	129,272	1,902	99%	202,789	107,504	65%

Annex 2: Glossary of terms

<i>Attributable fraction (AF)</i>	The proportion of cases that would not have occurred in the absence of occupational exposure.
<i>Attributable numbers (AN)</i>	Deaths, or cancer registrations, that would not have occurred in the absence of occupational exposure.
<i>Best source /study</i>	This is the study chosen as the source of a relative risk estimate for one or more exposure pairings.
<i>Cancer/exposure pairing</i>	This is a single carcinogenic occupational exposure and a site at which cancer is induced. It is the basic unit for which AF is calculated.
<i>Causality</i>	IARC criteria for causality include (1) a strong association (e.g. a large relative risk), (2) risk increasing with the exposure, and (3) the demonstration of a decline in risk after cessation, or reduction in exposure.
<i>Confounder</i>	A confounding factor, also known as an adjustment factor, is a factor which can have the effect of biasing the estimate of relative risk (RR) if it is correlated in the study population with the exposure of interest, whether it is in the causal pathway for the disease or not. Examples are age, and smoking status in the case of lung cancer.
<i>Dominant exposure</i>	Where exposures overlap in the working population, the exposed workers are allocated to the exposure deemed to have the highest risk, that is the dominant exposure, for calculation of the attributable fraction.
<i>“Established” exposure group</i>	Carcinogens established by IARC as Group 1 and with ‘strong’ evidence of carcinogenicity in humans, such as a large relative risk, for the relevant cancer site (after Siemiatycki et al (2004))
<i>“Uncertain” exposure group</i>	Carcinogens in IARC Groups 1 or 2A and with ‘suggestive’ evidence of carcinogenicity in humans for the relevant cancer site (after Siemiatycki et al., (2004)), plus other suspected carcinogens for which the evidence is not yet established.
<i>Exposure map</i>	A diagram setting out links and therefore overlaps between exposures in the working population. Linked exposures may also not act independently of one another in causing disease.
<i>Exposure scenario</i>	The industry process, job or occupation in which a worker is exposed to a carcinogenic substance.

<i>Exposure set</i>	A group of exposures that overlap in the working population, and are not independent of one another in causing disease, so that their contribution towards attributable fraction cannot be estimated separately.
<i>External data sources</i>	These are sources for an estimate of the proportion in the population exposed that do not come from the study providing the source for the relative risk estimate; either CAREX, national employment data from the Census of Employment (CoE) (and its successors the Annual Employment Survey (AES) and the Annual Business Inquiry (ABI)), or the Labour Force Survey (LFS).
<i>Industry-based study</i>	An epidemiological study whose study members are drawn from a single industry or occupational group. These are usually cohort studies, or case-control studies drawn from within an industry-based cohort.
<i>Occupational exposure</i>	There are two ways of defining exposure, by substance or by occupation. Substances are the IARC agents, groups of agents and mixtures identified as carcinogens. Occupations cover exposure scenarios where the causal agent is unknown.
<i>Overlapping exposures</i>	This is where the distribution of occupational exposures in the population overlaps, so that some workers will be subject to multiple exposures in a single workplace, or in a series of jobs over a working lifetime.
<i>Population-based study</i>	An epidemiological study drawing all study members from a national or large regional population, or from hospital attendees. Census based cohorts as well as hospital or cancer registry or death certificate based case-control studies are included under this heading.
<i>Relevant Exposure Period</i>	The period during which exposure to a causal factor is able to result in a cancer appearing at a certain time. It can also be called a Relevant Latency Window.
<i>Source population</i>	The population in which the study was done from which RR estimates for calculating AF have been taken.
<i>Target population</i>	The population of Great Britain

8. BIBLIOGRAPHY

- Andersen A, Barlow L, Engeland A, Kjaerheim K, Lynge E, Pukkala E (1999) Work-related cancer in the Nordic countries. *Scand J Work Environ Health* 25 (Suppl 2): 1–116.
- Benichou, J. and Gail, M.H., 1990. Variance calculations and confidence intervals for estimation of the attributable risk based on logistic models. *Biometrics*, 46: 991-1003.
- Benichou, J., 2001. A review of adjusted estimators of attributable risk. *Statistical Methods in Medical Research*, 10: 195-216
- Blanc PD, Becker CE. Occupational health planning in Israel: a US perspective. *Isr J Med Sci* 1992;**28**:500–8.
- Bouchardy, C., Schuler, G., Minder, C., Hotz, P. Bousquet, A., Levi, F., Fisch, T., Torhorst, J and Raymond, L., 2002. Cancer risk by occupation and socioeconomic group among men – a study by the Association of Swiss Cancer Registries. *Scand. J Work Environ Health*, 28, suppl 1: 1-88.
- Breslow, N. E. and Day, N. E., 1989. *Statistical Methods in Cancer Research, Volume II: The Design and Analysis of Cohort Studies*. P 67
- Bruzzi, P., Green, S.B., Byar, D.P., Brinton, L.A. and Schairer, C., 1985. Estimating the population attributable risk for multiple risk factors using case-control data. *American J Epidemiology*, 122, 904-914.
- Carpenter, L. and Roman, E., 1999. Cancer and Occupation in women: identifying associations using routinely collected national data. *Env. Health Perspectives*, 107: 299-303)
- Coggon, D., Harris, E.C., Poole, J. & Palmer, K.T. (2003). Extended follow-up of a cohort of British chemical workers exposed to formaldehyde. *Journal of the National Cancer Institute*, **95**, 1608-1615.
- Darnton, A.J., McElvenny, D.M. and Hodgson, J.T., 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. and Laird, N., 1986. Meta-analysis in clinical trials. *Controlled Clinical Trials*; 7: 177-188.
- Deschamps, F., Barouh, M., Deslee, G, Prevost, A., and Munck, J., 2006. Estimates of work-related cancers in workers exposed to carcinogens. *Occupational Medicine* 2006, 56: 204–209
- Doll R, Peto R. The causes of cancer: quantitative estimates of avoidable risks of cancer in the United States today. *J Natl Cancer Inst.* 1981;66(6):1191-1308.
- Dreyer, L., Andersen, A. and Pukkala, E., 1997. Occupation. In ‘Avoidable cancers in the Nordic countries’ *APMIS Suppl.* 76: 105, 68-79.
- Driscoll T, Mannetje A, Dryson E, et al. *The Burden of Occupational Disease and Injury in New Zealand: Technical Report*. Wellington: National Occupational Health and Safety Advisory Committee; 2004.
- Greenland S, Bias in methods for deriving standardized morbidity ratio and attributable fraction estimates, *Stat Med* 1984;3:131-41.
- Greenland S, Variance estimators for attributable fraction estimates consistent in both large strata and sparse data, *Stat Med* 1987; 6: 701-708
- Greenland, S.(1998), Meta-analysis. In: Rothman, K.J. & Greenland, S., eds, *Modern Epidemiology*, Philadelphia, Lippincott Williams & Wilkins, pp. 643–673

- Greenland, S., 2004. Interval estimation by simulation as an alternative to and extension of confidence intervals. *International J Epidemiology*, 33: 1389-1397.
- Gustavsson, P., Jakobsson, R., Nyberg, F., Pershagen, G., Jarup, L. and Scheele, P., 2000. Occupational exposure and lung cancer risk: A population-based case-referent study in Sweden. *American Journal Of Epidemiology*, 152, 32-40.
- Hodgson, J.T. and Darnton, A., 2000. The quantitative risks of mesothelioma and lung cancer in relation to asbestos exposure. *Ann. Occ. Hygiene*, 44, 565-601.
- Hoffmann K., Heidemann C., Weikert C., Schulze M.B., and Boeing H., 2006. Estimating the proportion of disease due to classes of sufficient causes. *Am J Epidemiol*;163: 76–83
- HSE (2002) Amendment to the Control of Asbestos at Work Regulations 1987 and ACoP: Regulatory Impact Assessment (available on the web at <http://www.hse.gov.uk/ria/chemical/ria-176.pdf>). Health and Safety Executive. P31 para A63.
- Kogevinas M, Kauppinen T, Boffetta P, Saracci R. *Estimation of the Burden of Occupational Cancer in Europe. Final Report to the European Commission of a project funded by the Programme "Europe Against Cancer"*. Barcelona: IMIM, 1998.
- Landrigan, P.L., Schechter, C.B., Lipton, J.M., Fahs, M.C., and Schwartz, J., 2002. Environmental pollutants and disease in American children: estimates of morbidity, mortality and costs for lead poisoning, asthma, cancer, and developmental disabilities. *Environmental Health Perspectives*, 110, 721-728.
- Leigh, J., Markowitz, S and Fahs, M., 1997. Occupational injury and illness in the United States. Estimates of costs, morbidity and mortality. *Arch. Internal Med.*, 157: 1557-1568.
- Levin M, The occurrence of lung cancer in man, *Acta Unio Internationalis Contra Cancrum* 1953; 9:531-541
- Luce, D., Leclerc, A., Begin, D., Demers, P.A., Gerin, M., Orłowski, E., Kogevinas, M., Belli, S., Bugel, I., Bolm-Audorf, U., Brinton, L.A., Comba, P., Hardell, L., Hayes, R.B., Magnani, C., Merler, E., Preston-Martin, S., Vaughan, T.L., Zheng, W. & Boffetta, P. (2002). Sinonasal cancer and occupational exposures: a pooled analysis of 12 case-control studies. *Cancer Causes and Control*, 13, 147-157.
- Mannetje, A., Kogevinas, M., Luce, D., Demers, P.A., Begin, D., Bolm-Audorf, U., Comba, P., Gerin, M., Hardell, L., Hayes, R.B., Leclerc, A., Magnani, C., Merler, E., Tobias, A. & Boffetta, P. (1999). Sinonasal cancer, occupation, and tobacco smoking in European women and men. *American Journal of Industrial Medicine*, 36, 101-107.
- Markowitz SB, Fischer E, Fahs MC. Occupational disease in New York state: a comprehensive examination. *Am J Ind Med* 1989; 16: 417–35.
- Mathers CD, Vos ET, Stevenson CE, Begg SJ. The Australian burden of disease study: measuring the loss of health from diseases, injuries and risk factors. *Med J Aust*. 2000; 172:592–596
- Miettinen O: Proportion of disease caused or prevented by a given exposure, trait or Intervention, *Am J Epidemiology* 1974;99:325-32
- Morabia A, Markowitz S, Garibaldi K, Wynder EL. Lung cancer and occupation: results of a multicentre case– control study. *Br J Ind Med*. 1992;49: 721–727.
- Morrell, S., Kerr, C., Driscoll, T., Taylor, R., Salkeld, G. and Corbett, S., 1998. Best estimate of the magnitude of mortality due to occupational exposure to hazardous substances. *Occup. Environ. Medicine*, 55:634-641.

- Nelson, D.I., Concha-Barrientos, M., Driscoll, T., Steenland, K., Fingerhut, M., Punnett, L., Pruss-Ustun, A., Leigh, J and Corvalan, C., 2005. The global burden of selected occupational diseases and injury risks: methodology and summary. *American J Ind. Med.*, 48: 400-418.
- Nurminen M, Karjalainen A. Epidemiologic estimate of the proportion of fatalities related to occupational factors in Finland. *Scand J Work Environ Health*. 2001;27:161–213.
- ONS, 2005. Mortality statistics. Series DH2 no. 30 Table 2.2
- ONS, 2006. Cancer statistics, Registrations. Series MB1 no. 34 Table 1
- OPCS, HSE, Drever, F (ed.), 1995. Occupational Health Decennial Supplement. The Registrar General's decennial supplement for England and Wales. Series DS No. 10. London, HMSO.
- OPCS, HSE, Drever, F. (1995). *Occupational Health Decennial Supplement*. Series DS no. 10: The Registrar General's Decennial Supplement for England and Wales. HMSO: London.
- Rockhill B, Newman B, Weinberg C, Use and misuse of population attributable fractions, *Am J Public Health* 1998;88:15-9.
- Rothman, K.J and Greenland, S., 1998. *Modern Epidemiology*. P. 660.
- Siemiatycki, J., Richardson, L., Straif, K., Latreille, B., Lakhani, R., Campbell, S., Rousseau, M-C. and Boffetta, P., 2004. Listing Occupational Carcinogens. *Environmental Health Perspectives* 112, 1447 – 1460.
- Steenland K, Burnett C, Lalich N, Ward E, Hurrell J. Dying for work: the magnitude of US mortality from selected causes of death associated with occupation. *Am J Ind Med*. 2003;43:461– 482.
- Steenland, K. and Armstrong, B., 2006. An overview of methods for calculating the burden of disease due to specific risk factors. *Epidemiology*, 17 (5): 512-519.
- Steenland, K., Loomis, D., Chy, C. & Simonsen, N. (1996). Review of occupational lung carcinogens. *American Journal of Industrial Medicine*, **29**, 474-490.
- vanLoon, AJM., Kant, IJ., Swaen, GMH., Goldbohm, RA., Kremer, AM. and vandenBrandt, PA., 1997. Occupational exposure to carcinogens and risk of lung cancer: results from The Netherlands cohort study. *Occupational And Environmental Medicine*, 54, 817-824.
- Vineis P, Simonato L. Proportion of lung and bladder cancers in males resulting from occupation: a systematic approach. *Arch Environ Health*. 1991; 46:6 –15.
- Vineis, P., Thomas, T., Hayes, R.B., Blot, W.J., Mason, T.J., Pickle, L.W., Correa, P., Fontham, E.T.H., and Schoenberg, J., 1988. Proportion of lung cancers in males, due to occupation, in different areas of the USA. *Int. J. Cancer*, 42, 851-856.

The burden of occupational cancer in Great Britain

Methodology

The aim of this project was to produce an updated estimate of the current burden of occupational cancer specifically for Great Britain. The primary measure of the burden of cancer used was the attributable fraction (AF), ie the proportion of cases that would not have occurred in the absence of exposure. Data on the risk of the disease due to the exposures of interest, taking into account confounding factors and overlapping exposures, were combined with data on the proportion of the target population exposed over the period in which relevant exposure occurred. Estimation was carried out for carcinogenic agents or exposure circumstances that were classified by the International Agency for Research on Cancer (IARC) as Group 1 or 2A carcinogens with strong or suggestive human evidence. Estimation was carried out for 2004 for mortality and 2003 for cancer incidence for cancer of the bladder, leukaemia, cancer of the lung, mesothelioma, non-melanoma skin cancer (NMSC), and sinonasal cancer.

The proportion of cancer deaths in 2004 attributable to occupation was estimated to be 8.0% in men and 1.5% in women with an overall estimate of 4.9% for men plus women. Estimated numbers of deaths attributable to occupation were 6,259 for men and 1,058 for women giving a total of 7,317. The total number of cancer registrations in 2003 attributable to occupational causes was 13,338 for men plus women. Asbestos contributed the largest numbers of deaths and registrations (mesothelioma and lung cancer), followed by mineral oils (mainly NMSC), solar radiation (NMSC), silica (lung cancer) and diesel engine exhaust (lung and bladder cancer). Large numbers of workers were potentially exposed to several carcinogenic agents over the risk exposure periods, particularly in the construction industry, as farmers or as other agricultural workers, and as workers in manufacture of machinery and other equipment, manufacture of wood products, land transport, metal working, painting, welding and textiles. There are several sources of uncertainty in the estimates, including exclusion of other potential carcinogenic agents, potentially inaccurate or approximate data and methodological issues. On balance, the estimates are likely to be a conservative estimate of the true risk. Future work will address estimation for the remaining cancers that have yet to be examined, together with development of methodology for predicting future estimates of the occupational cancers due to more recent exposures.

This report and the work it describes were funded by the Health and Safety Executive (HSE). Its contents, including any opinions and/or conclusions expressed, are those of the authors alone and do not necessarily reflect HSE policy.