

A review of the data quality and comparability of case–control studies of low-level exposure to benzene in the petroleum industry

B. G. Miller · W. Fransman · D. Heederik ·
J. F. Hurley · H. Kromhout · E. Fitzsimons

Received: 1 December 2008 / Accepted: 24 August 2009 / Published online: 16 September 2009
© Springer-Verlag 2009

Abstract

Purpose Published case–control studies of risks of leukaemia following low exposures to benzene in the distribution of petroleum (gasoline) have not all identified the same level of risk, but the studies have had differences in cohort inclusion, case determination and availability of occupational and lifestyle data. We reviewed the quality and comparability of the data from three (of four) studies.

Methods Through site visits, discussions with the investigators and reading study reports, we reviewed and audited the methods used for selecting cases and controls, for estimating individual exposures and for analysing and interpreting the data. Case–control comparisons of exposures were examined using customised graphs.

Results We found that

- there were no issues of subject selection, methods or general data quality that were likely to have distorted their internal comparisons;
- we could not check in detail whether the metric for exposure assessments was the same across the studies;
- the exposure assessments for the Australian study required the least backward estimation, and the Canadian, which also had fewest cases, the most;

- evidence of an increased risk at higher exposures in Australia was convincing.

Conclusions The findings are consistent with some effect of benzene at higher lifetime exposures. A proposed pooled analysis should improve quantification of any exposure–response relationship.

Keywords Petroleum · Benzene · Leukaemia

Introduction

Benzene is an aromatic hydrocarbon that is used as a solvent and as a chemical feedstock in the production of dyes, pharmaceuticals, synthetic rubber, nylon and pesticides. Benzene has been a component of automobile fuel (‘petrol’: US ‘gasoline’) at levels that have varied by region, crude source and era: since the 1950s, benzene concentrations in petrol have generally declined overall. Apart from occupational exposure, benzene exposure may derive from combustion of domestic fuels and tobacco.

The International Agency for Research on Cancer classified benzene as carcinogenic to man, in 1982 (IARC 2004). Plausible pathways and mechanisms for benzene’s mode of action have been proposed, and there is now a consensus that high exposures to benzene bring an increased risk of leukaemia and in particular to acute myeloid leukaemia (Rinsky et al. 1987). However, the question remains whether leukaemia risks are elevated at lower exposures.

There have been a number of cohort studies of workers involved in the distribution and marketing sectors of the petrol industry, such as tanker drivers and loaders, forecourt attendants and others who were likely to have had exposures to benzene at lower levels than in some other

B. G. Miller (✉) · W. Fransman · J. F. Hurley
Institute of Occupational Medicine, Research Avenue North,
Riccarton, Edinburgh EH14 4AP, UK
e-mail: brian.miller@iom-world.org

W. Fransman · D. Heederik · H. Kromhout
Institute for Risk Assessment Sciences, University of Utrecht,
Utrecht, The Netherlands

E. Fitzsimons
Department of Haematology,
University of Glasgow, Glasgow, UK

industries. Cohort studies have typically had relatively crude exposure characterisations, and interpretation of SMRs has been complicated by healthy worker effects. More detailed nested case–control studies have attempted to address the question of risks of benzene at a greater level of detail, focussing on lymphatic–haematopoietic cancers as an outcome. Principal among these have been the studies of workers from the petroleum marketing and distribution workers from Canada (Schnatter et al. 1996), UK (Rushton and Romaniuk 1997), USA (Wong et al. 1999) and Australia (Glass et al. 2003). The results from the first three studies were broadly similar in showing little evidence of an effect of low benzene exposure. However, the Australian study reported a leukaemogenic effect at much lower levels than the other studies, down to cumulative exposures >2 ppm-years. Schnatter (2004) suggested that this finding could be due to an unusually low number of cases in the lowest (reference) quintile of exposure, inflating odds ratios for the other quintiles. Doll (Institute of Petroleum 2003) suggested that the high relative risks reported ‘could not be regarded as reflecting causality’ because they were not consistent with the results from the cohort study. Goldsmith (2004) suggested that blood testing in workers more heavily exposed to benzene might have introduced surveillance bias, increasing the chances of detecting cases in that group, although this was refuted by the authors (Glass et al. 2004); also that evidence for an increase at low exposures in incidence of acute myeloid leukaemia, the type most commonly associated with benzene exposure, was absent. Comparisons were made with more difficulty because each study used slightly different methods and procedures, and the possibilities for these to produce the different results were not clear.

This collaborative study was set up to evaluate data consistency and quality in the existing case–control studies of leukaemia in oil distribution and refinery workers, to indicate the relevance of any difference for the findings of the studies and to discuss aspects important for the possible pooling of the studies, to gain power. The primary purpose of our work was to elucidate the reasons for inconsistencies by characterising the similarities and differences among the four studies. An initial assessment revealed that it would not be possible to arrange access to materials from the study of US workers, and this review has therefore investigated in detail only three studies, of workers in Australia, Canada and the United Kingdom.

Methods

It was agreed with a Scientific Steering Committee (SSC) that the audit should focus on the following questions:

- differences in definition of cases;
- differences in methods of diagnosis underlying cause-of-death certification or cancer registration;
- differences in power to detect small effects and/or effects at low exposures, from either considerations of study size or statistical analysis techniques;
- differences in the level of detail available to draw job and/or task distinctions in individual work histories;
- differences in the levels of exposure assigned in the exposure assessment exercise;
- differences in the adjustment for external confounding factors.

We created and agreed with the SSC and the studies’ principal investigators (PIs), a protocol for site visits consisting of descriptive text summarising the background and general approach, a list of 64 specific questions and a draft structure for each visit, to be tailored to local circumstances. The questions covered epidemiological aspects of study design; exposure assessment and assignment; estimation of individual exposures; statistical analyses; and validation and reliability.

Two of the audit team, supplying expertise in epidemiology and exposure assessment, visited a single central point for each study, to inspect and assess the available records and documentation in detail. The records for the Australian cohort study and its nested case–control study were held separately in Adelaide and Melbourne, so visits to both sites were necessary.

After a preparatory meeting with the PIs from the three studies to discuss the protocol and its implications, the audit team were provided with background documentation from each study, including procedural handbooks and detailed final reports.

After these documents had been studied, the site visits took place in the summer of 2004. At each site, the audit team consulted with the PI and with other team members who were available for interview. The interviews focussed on the questions in the protocol. During these visits, we also obtained, for some 5–10% of cases and controls selected at random as a representative sample from each study, fully anonymised copies of the occupational history data, and the computer records held were checked against these.

Detailed site reports on the findings were shown to the study teams to check for accuracy, and revised as necessary, and a final summary report was submitted to the sponsors (Miller et al. 2005).

In what follows we refer to the different studies by the following abbreviations:

- IOL: study on Canadian employees of Imperial Oil Ltd. and associated companies

Table 1 Comparative characteristics of parent cohort studies

	IOL	IP	HW
Cohort study			
Industry segments	Marketing/distribution	UK oil distribution centres	Marketing, distribution, upstream, refining
Number of sites	226	315	235
Source of ID records	Computerised employee relations database	Personnel and pensions records	Company records
Inclusion criteria	All active employees and annuitants 01/01/1964 All new regular employees hired 01/01/1964–31/12/1983	Employed 1+ year 01/01/1950–31/12/1975	Employed 5+ years since 1980
Exclusions	No females	No females	
Cohort size	6,672	23,306	16,252 male, 1,273 female
Follow-up period	01/01/1964–31/12/1983	01/01/1950–31/12/1975	01/01/1980–31/12/1998 (deaths) 31/12/1996 (incidents)
Number of deaths	1,154	8,743	883 + 520 incident cancers

- IP: UK study of workers from four companies affiliated to the Institute of Petroleum
- HW: Australian Health Watch study

Results

Table 1 lists many detailed aspects of the cohort studies, and Table 2 lists those of their nested case–control studies.

For the case–control studies, the study populations were entirely male and were all from the same industry: the IOL and IP cohort studies contained a rather higher proportion of individuals in jobs where exposure to benzene was assessed at a background level compared with the HW study, and this is reflected in the exposure distribution of the subjects.

Cases in the IOL study were identified from the cohort study data files containing the results of the tracing exercises, yielding, between 1964 and 1983, 16 leukaemias, seven multiple myelomas and eight non-Hodgkin lymphomas. This study did not identify the particular types (acute or chronic, myeloid or lymphoid) of leukaemia found in these 16 cases. With such a small total number of cases, analyses of subtypes would have had very low power.

Cases in the IP study were defined as deaths or diagnosed cases of leukaemia only, traced through the UK's death registration or cancer registration systems via the National Health Service Central Register. A total of 91 leukaemia cases were identified in the IP study, and the different types were distinguished in the analyses: acute leukaemia 42 cases; acute myeloid leukaemia (AML) 32; acute lymphoblastic leukaemia (ALL) 7 and not specified 3. There were 43 cases of chronic leukaemia: chronic myeloid leukaemia (CML) 11, chronic lymphatic leukaemia (CLL)

31 and chronic monocytic leukaemia 1. A further 6 cases were not adequately typed to allow further classification.

The HW case–control study defined the outcome as a “newly diagnosed lympho-haematopoietic cancer” and was thus an incidence study. For ethical reasons, each case had to be reported to Health Watch either by himself or by his family, and was confirmed by pathology report, cancer registration, letter from medical practitioner or death certificate. The researchers were ethically precluded from verifying case status unless the case was already dead or lost to follow-up in the cohort study. Only one otherwise eligible case traced through the cancer registry had to be omitted because he had not self-reported. HW identified 79 cases, of which 33 were leukaemias. Medical information for the 9 cases where diagnosis was uncertain was reviewed by an expert haematologist, confirming or revising the type of leukaemia involved (AML 11, ALL 2, CLL 11, CML 6 and other lymphoid leukaemias 6).

The HW study required the voluntary participation of its recruits, because data collection included questionnaires to the cohort members. This is unlikely to have introduced bias in the study composition, because co-operation by subjects requested to participate, in the early days when most of the subjects were recruited, was well in excess of 90%.

In all the studies, the outcomes were classified according to either the 8th or the 9th revision of the International Classification of Diseases (ICD). For leukaemia classification, the different revisions are the same, and for discrimination between leukaemia and other diseases, it is immaterial whether 8th or 9th ICD revision is employed. The definition of casehood differed between studies, in that the IOL study included only deaths, while the IP and HW studies included deaths and cancer registrations. (The IP study had death certificates on 88 of their 91 cases.)

Table 2 Comparative characteristics of nested case–control studies

	IOL		IP	HW
Case-control study				
Case definition	Died from leukaemia and ever worked in marketing, distribution, marine or pipeline segments		Died before 01/01/1993 from leukaemia OR cancer registration with leukaemia	Male with reported newly diagnosed lympho-haematopoietic cancer, confirmed by certificate.
Cases of leukaemia	16		91	33
Other blood cancers ^a	7 MM, 8 NHL		Not recorded	15 MM, 31 NHL
Case type(s)	Decedent		Decedent and incident	Decedent and incident
Matching criteria	Decade of birth, alive at case identification		Company, age, alive at case identification	Year of birth, alive at case death/registration
#Controls per case	4		4	5
Leukaemia subtypes analysed	ALL, CLL, AML, CML		ALL, CLL, AML, CML	ALL, CLL, AML, CML
Source of work histories	Electronic & hard copy personnel records		Personnel records, pension fund, medical records, interviews	Computerised job history, interview of case or proxy
Benzene exposures considered ^b	CE (ppm-years), mean ppm, peak intensity, dermal exposure potential		CE (ppm-years), mean ppm, peaks, dermal exposure	CE (ppm-years), mean ppm, duration,
Lags considered	0, 5, 10, 15 years		0, 5, 10 years	0, 5, 10, 15 years
Confounders adjusted for ^c	Smoking, SE type, chest X-rays		Employment status at follow-up end, SE type, start date, previous driving jobs	Start date, smoking, alcohol, country of birth ^a

^a MM multiple myeloma, NHL non-Hodgkin lymphoma^b CE cumulative exposure^c SE socioeconomic

The date of definition for a case registration is necessarily earlier than if the case had been identified only at death, which in turn affects the age and exposure distribution of potential controls. Analyses of the data have not to date taken account of this difference, and it is not easy to see how a satisfactory allowance could be made, since the distribution of times between the unknown diagnosis dates and the dates of deaths is likely to have been very variable. However, the number of cases identified from cancer registration alone was very small: although 30–50% of AMLs may be cured or placed in remission by modern treatments, the proportion was much smaller in the past, and secondary acute leukaemia following chemical exposure has a much poorer prognosis than *de novo* cases. We judged that none of the studies was likely to have missed identifying significant numbers of cases.

In all the studies, the controls were same-sex. For the IOL study, controls were selected with replacement from the cohort database at a ratio of 4:1, matched by decade of birth and alive on the case's date of death. In the IP study, controls were selected without replacement at a ratio of 4:1, from the same company and with the year of birth within 3 years of the case. In the HW study, controls were selected from the cohort study database at a ratio of 5:1, matched on sex and year of birth, sampled with replacement. Different case-control ratios are not a source of bias, and the difference in power to detect a relationship between a 4:1 and a 5:1 ratio is small. These slight differences in the selection of controls were not likely, in our opinion, to have had any serious effect on the study results.

All the study teams took great pains to compile the best possible work histories for the case control studies, working with what records were available. The IOL study got its work histories direct from the company personnel records, and these were relatively complete. The IP study also used company personnel records, but there were some missing data, and some supplementation from other sources, including pension records, medical records and interviews with retired or long service staff. In the HW study, work histories were given by individuals at interview on recruitment to the cohort study, and later cross-checked with company records. In all of these cases, the histories were based or verified on data that predated the case definitions, and thus would not have been subject to a leukaemia-related reporting bias. We found nothing in any of the studies to suggest systematic differences between the reliability or completeness of the work histories of cases and their matched controls.

In the IOL study, other data collected from company medical records included information on smoking habits, hobbies, previous occupations and exposures, diagnostic radiation exposures and family history of cancer. These data were not available in all cases, but those that were gave a fair amount of detail. In the IP study, data available

from medical records were extracted on smoking histories and on previous occupations. Smoking data were available for only a small minority and were unknown for almost 90% of cases and controls. For the HW study, the cohort database included data from the health survey questionnaires on smoking habits, typical alcohol consumption, and previous employments. The smoking data were relatively detailed, including amounts smoked.

The retrospective exposure assessment methodologies for the three studies used an approach designed for the IOL study (Armstrong et al. 1996). Exposure assessment started from "base estimates" of arithmetic mean exposures (in ppm), based on measurements for typical jobs and periods. These were subsequently adjusted for modifying factors (K-factors) assigned by experts, to allow for differences in local conditions and practices, to result in a time-weighted workplace exposure estimate for each line (episode) of a subject's work history. These were multiplied by the time spent in that task/job and summed to result in a cumulative exposure estimate in ppm-years. The extent of extrapolation and the use of modifying factors differed considerably (Miller et al. 2005), and we cannot exclude the possibility of systematic differences in the resulting exposure estimates. It is likely that the retrospective exposure estimates of the HW study were more accurate and reliable than those of the IOL and IP studies, because they were based on more recent measurements.

A relatively large group of subjects in the IOL study was exposed to very low levels of benzene (41% of the population had only background exposure), and only a few subjects were exposed to high levels of benzene (15% of all subjects had spent the majority of their time in jobs with exposure to benzene) compared to the other two case-control studies. The HW study consisted of mainly exposed workers, but its younger cohort did not have particularly high exposures to benzene (maximum was 50.9 ppm-years). The IOL and IP studies included some individuals with very high exposures from the earlier periods in the cohort (pre-1940).

We judge that the exposure assessment approach used in all three studies will have led to an accurate ranking of subjects within each study. Differences in input data, application of modifying factors (K-factors) and thus the extent of extrapolation from the base estimate situation to earlier periods, may have led to some systematic differences between studies in the levels assigned to comparable jobs.

The statistical analysis of case-control studies is based on comparing the attributes of cases with those of comparable controls. Where matching is used, the analysis needs to respect the matching, else the estimates of relative risk may be biased. All the studies used the appropriate conditional regression techniques in their analyses, each fitting numerous models with different combinations of predictors.

Analyses detailing the different types of leukaemia were carried out only in the IP and HW studies, but were inconclusive, because of the small numbers of cases once subdivided. With modern diagnostic techniques, it seems unlikely that there would be much disagreement on distinctions between acute lymphoid and acute myeloid conditions, but changes in classifications of just a few cases could have a relatively large impact on the study results. One case in the IP study was mistakenly classified as AML instead of CML.

The principal exposure measure for each study was an estimated lifetime cumulative exposure (CE) in ppm-years. In some models, exposure was included as a continuous variable, but more often the exposures were aggregated into groups and relative risks calculated against the lowest-exposed group, treated as a baseline. Lags of 0, 5 and 10 years were fitted, and 15 years in IOL and HW. Some analyses also used average or peak ppm and estimates of potential dermal exposure.

One potentially confounding variable was tobacco smoking, and the availability of smoking data varied greatly. The questionnaires used in the HW cohort study gave smoking habits for almost all subjects. In the IOL study, 7 of 16 cases had smoking data. In the IP study, smoking habits were not known for most of the subjects. However, in the HW study, risk of leukaemia was not found to be associated with smoking habits, suggesting that any link of leukaemia with smoking may be weak, in that case, its omission as an explanatory variable may have little or no effect. Rushton and Romaniuk (1997) noted that reviews of the topic have suggested, at most, a mild increase in risk for smokers, in accordance with current clinical views. Other variables included as predictors varied between the studies: socioeconomic status in IOL and IP, employment start date in IP and HW. (See Table 2) In principle, any of these variables could have introduced differences in the results, but overall, none of these was judged likely to be an important confounder.

It was possible to compare the derived exposure–response results from each of the studies only in qualitative terms, because the exposure categories had been defined differently. Since the risks in higher exposure categories were described relative to these different baselines, some of which were relatively poorly determined, it was difficult to compare the levels of relative risk across studies. This is a strong argument for deriving combined estimates by analysing pooled data and generating a common baseline, rather than by a meta-analysis of published categorical relative risks.

A particular case in point is the HW study. Their results (Glass et al. 2003; Table 4) showed that the four exposure groups between 1 and 16 ppm-years, compared to a baseline group with exposures up to 1 ppm-year, had odds ratios

of 3.9, 6.1, 2.4 and 5.9. The two highest of these were declared statistically significant at 5%, although there was no obvious trend with exposure. In addition, the highest group, with exposures greater than 16 ppm-years, had an odds ratio of 98.2, which was highly significant. Different choices of group limits alter the judgment on whether there are significantly increased risks in the middle exposure categories, but the finding of an increased risk in the highest exposures is robust to these redefinitions (Institute of Petroleum 2003). However, standard reference rates applied to the cohort data showed that the cases in the baseline group were fewer than expected, suggesting that the reference level was artificially low, introducing a small-sample bias in the other estimates (Greenland et al. 2000); this could easily have arisen by chance. A synthesis of these results is that there is little convincing evidence of any increase in risk below 16 ppm-years, but that it is plausible that exposures above this level carry an increased leukaemia risk.

Matching in case–control studies complicates the graphical presentation of data, and studies are often presented without graphical displays, or graphed with heavily grouped exposure categories, which can make it hard to compare across studies. However, within each matched set, it is the difference between the exposures of the case and the matched set of controls that is important, and the simplest option is to plot the distribution of those differences, possibly on the logarithmic scale (i.e. as ratios), perhaps as a histogram, probability plot or box-and-whisker summary.

Figure 1 summarises case–control comparisons for the three studies. In each graph, the x-axis is the geometric mean exposure of the whole case–control set, log-scaled, and these are summarised by a box-and-whisker plot above the graph frame. The box stretches from the 25th to the 75th percentile with a central line for the median; the whiskers go out to the 10th and 90th percentile; and the highest 10% and lowest 10% of values are plotted individually. Box widths are scaled to the square roots of the numbers of case–control sets. In the HW study, average exposures overall tended to be higher than in the other studies: a higher proportion of case–control means lay above 1 ppm-year.

To the right of each graph is a box-and-whisker plot summarising the distribution of the case–control exposure ratios, again log-scaled. If exposure had no influence on risk, these summaries would be distributed about unity. The IP ratios are roughly symmetric around a case–control ratio of 1: the IOL study is sparse, but does not suggest a shift from symmetry. In contrast, for the HW study, the distribution's centre is clearly shifted away from 1, suggesting an exposure effect.

In the two-dimensional scatter graphs, we plot each ratio against the corresponding geometric mean. In the absence of an effect of exposure, this mean will tend to reflect

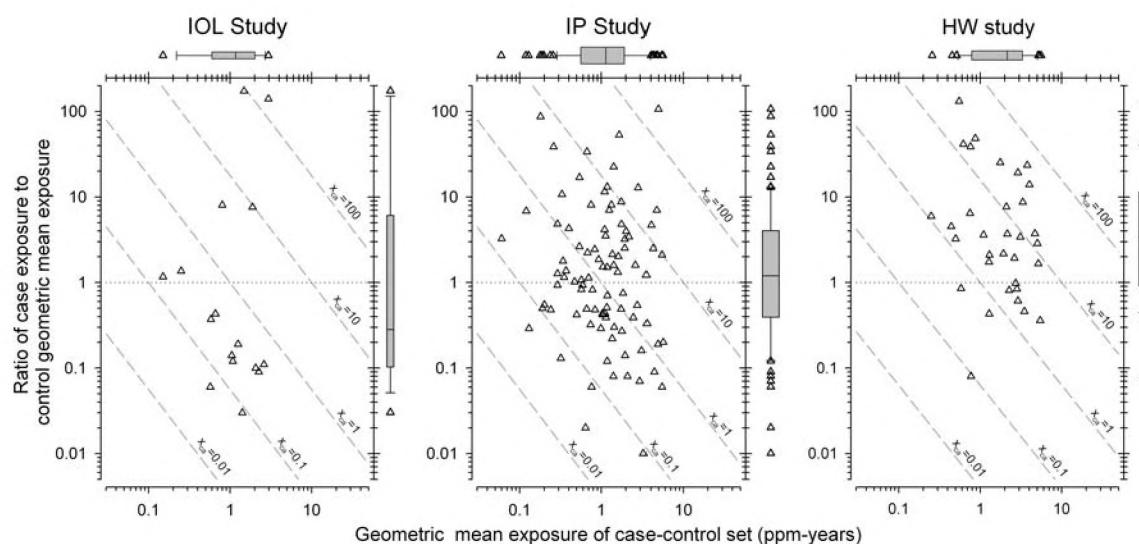


Fig. 1 Comparisons of exposures between cases and their matched controls, in box-plots and plotted against geometric mean of exposures (ppm-years) in case–control set

differences in exposure between the sets, which is likely to be correlated with the matching variable age. However, if there is an exposure effect, it will be driven by the exposures of cases. The dashed diagonal lines within the graphs correspond to constant values of the case exposure x_{ca} . From these, we observe that almost all cases in the HW study had exposures above 1 ppm-year, and a much higher proportion above 10 ppm-years than in the IP study.

These graphs were designed specifically for this study, and may be useful elsewhere in showing the range of exposures for case–control comparisons and in demonstrating the existence of an exposure–response relationship. However, they cannot display the shape of the underlying relationship: for this, we suggest fitting smoothed curvilinear models within a conditional regression framework.

Discussion

In carrying out this review, we received willing co-operation from the study teams, and access to whatever study materials we requested that they could find. Although the investigators from the original studies had co-operated closely in study design and development of methods (e.g. for exposure estimation), we found numerous detailed differences between the studies. Mostly, these were related to the circumstances of who was studied, and what data could be found for them, rather than differences in methods per se. We judged that those differences were not responsible—individually or together—for the apparent differences in results between the studies.

In the detail of the exposure assessments, it was clear that the data available on past benzene concentrations studies

covered different periods and therefore different degrees of extrapolation and assumption to cover periods without measurements. In addition, it was hard to judge to what extent the K-factors used for adjustments and extrapolations were comparable across studies, and we have to allow the possibility that the cumulative exposures calculated, while correctly ranking individuals within each study, may not scale exactly across the studies (Miller et al. 2005).

It was clear that the differences between studies led to differences in power to detect relationships between benzene and leukaemia. We judged that the apparent discrepancies in results were not due to confounding or due to inclusion or information biases; more likely, to study size and to an extent to differences in distributions of exposure. The IOL study was too small to be conclusive, and had predominantly low exposures. The HW study had the highest exposures on average, and showed results consistent (or not inconsistent) with increased leukaemia risks in the higher exposure ranges, e.g. with cumulative benzene exposures in excess of about 10 ppm-years. The apparent difference between the lowest and middle exposure groups in HW was almost certainly a small-sample bias caused by comparing with a baseline group with lower than expected mortality—a useful check in any nested case–control study.

Combining the data from the three studies would undoubtedly bring greater power. We recommended that a simple meta-analysis of the published data summaries was not desirable, because the exposure categories used to group the quantitative exposures differ; we recommended a combined analysis, an option considered by the study principals (Schnatter 2004).

The principal investigators of the three studies have now been commissioned to collaborate on a pooled study using

common exposure categories, investigating leukaemia cell types more thoroughly and making some standardisations across the exposure calculations, e.g. in assigning a common level for background concentrations. The complementary exposure distributions in Fig. 1 suggest that a pooled exposure distribution will be strengthened across the range. To maximise the number of cases, new cases and controls are to be recruited and their work histories collected. To add power to analyses of specific cell types, we recommended that the information on cell type for all cases should be reviewed centrally by one or more expert haematologists, to agree a final standardised classification. We also recommended that the expert opinions and modifying K-factors used in the exposure assessments should be revisited, and the factors validated and standardised across the studies. Finally, we suggested that statistical methods for a pooled analysis should be as little influenced by grouping strategies as possible, e.g. by fitting generalized additive models or other smoothed curves to the pattern of the data (Hastie and Tibshirani 1990).

Carrying out this review has shown that careful examination of original study data, with the active co-operation of the relevant research teams, may resolve apparent inconsistencies in study results, and prepare the ground for a pooled analysis of data.

Acknowledgments We are indebted to the principal investigators of the original studies, who gave help, co-operation and encouragement without which this review would have been impossible. Mark Nicolich of Exxon Mobil BSI spotted a mistake in an earlier version of Fig. 1, corrected in the current improved version. We thank CONCAWE for funding the work: the review benefited from advice from a Scientific Steering Committee appointed by them. We thank also our colleagues and three referees, all of whose comments improved this paper.

Conflict of interest statement The review project was funded by the petroleum industry association CONCAWE. The authors retained full and independent editorial control over all reporting.

References

- Armstrong TW, Pearlman ED, Schnatter AR et al (1996) Retrospective benzene and total hydrocarbon exposure assessment for a petroleum marketing and distribution worker epidemiology study. *Am Ind Hyg Assoc J* 57:333–343
- Glass DC, Gray CN, Jolley DJ et al (2003) Leukemia risk associated with low-level benzene exposure. *Epidemiology* 14:569–577
- Glass DC, Gray CN, Jolley DJ et al (2004) Benzene exposure and leukemia (Authors' Response to Letters). *Epidemiology* 15:510–511
- Goldsmith B (2004) Benzene exposure and leukemia (Letter to the Editor). *Epidemiology* 15:509–510
- Greenland S, Schwartzbaum JA, Finkle WD (2000) Problems due to small samples and sparse data in conditional logistic regression analysis. *Am J Epidemiol* 151(5):531–539
- Hastie TJ, Tibshirani RJ (1990) Generalized additive models. Chapman & Hall, London
- IARC (International Agency for Research on Cancer) (2004) IARC monographs on the evaluation of carcinogenic risks to humans, vol 82. IARC, Lyon
- Institute of Petroleum (2003) Leukaemia risks in relation to benzene exposure. Report of a meeting held on 22nd October 2002, London: Institute of Petroleum
- Miller BG, Fransman W, Heederik D et al. (2005) A review of the data quality and comparability of case-control studies of low-level exposure to benzene in the petroleum industry. Edinburgh: Institute of Occupational Medicine. (IOM Research Report TM/05/04). http://www.iom-world.org/pubs/IOM_TM0504.pdf
- Rinsky RA, Smith AB, Hornung R et al (1987) Benzene and leukaemia. An epidemiologic risk assessment. *N Engl J Med* 316:1044–1050
- Rushton L, Romaniuk H (1997) A case-control study to investigate the risk of leukaemia associated with exposure to benzene in petroleum marketing and distribution workers in the United Kingdom. *Occup Environ Med* 54:152–166
- Schnatter AR (2004) Benzene exposure and leukemia (Letter to the Editor). *Epidemiology* 15:509
- Schnatter AR, Armstrong TW, Nicolich MJ et al (1996) Lymphohematopoietic malignancies and quantitative estimates of exposure to benzene in Canadian petroleum distribution workers. *Occup Environ Med* 53:773–781
- Wong O, Trent L, Harris F (1999) Nested case-control study of leukaemia, multiple myeloma, and kidney cancer in a cohort of petroleum workers exposed to gasoline. *Occup Environ Med* 56:217–221