Institute of Occupational Medicine

Research Report TM/05/04
July 2005

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

BG Miller, W Fransman, D Heederik, JF Hurley,
H Kromhout and E Fitzsimons

Research Report

Research Report TM/05/04
July 2005

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

BG Miller[1], W Fransman[1,2], D Heederik[2], JF Hurley[1], H Kromhout[2] and E Fitzsimons[3]

[1]Institute of Occupational Medicine, Edinburgh
[2] Institute for Risk Assessment Sciences, University of Utrecht
[3] Department of Haematology, University of Glasgow

Three nested case-control studies (in Canada, UK and Australia) of risks of leukaemia following low exposures to benzene in the distribution of petroleum (gasoline) produced apparently inconsistent results: the Canadian study reported no excess risks at the exposure levels studied, the Australian study reported risks especially at higher exposures, and the UK study gave indeterminate results.

Through site visits, structured discussions with the investigators, and detailed reading of published and unpublished study reports, we reviewed and audited the methods used for selecting cases and controls, for estimating individual exposures, and for analysing and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. Where possible, we examined the data graphically.

We found that:

- all of the studies had been well performed; 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 whether the metric for exposure assessments is 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.

Given the relative strengths of the three studies, the findings are consistent with some effect of benzene at the high end of the subjects' lifetime exposures. An analysis pooling data from all three studies would improve quantification of any exposure-response relationship, but would benefit from preliminary work to demonstrate or improve comparability of the exposure assessments and of cell typing in the cases.

INSTITUTE OF OCCUPATIONAL MEDICINE
Research Park North, Riccarton, Edinburgh, EH14 4AP
Tel: +44 (0)870 850 5131 Fax: +44 (0)870 850 5132
e-mail publications@iomhq.org.uk

CONTENTS

1 INTRODUCTION

1.1 BACKGROUND

Benzene is an aromatic hydrocarbon, obtained principally by catalytic cracking and fractional distillation of organic feedstocks such as coal and crude oil. From the mid-nineteenth century it found wide use as a solvent, and as a chemical feedstock in the production of dyes, pharmaceuticals, synthetic rubber, nylon and pesticides. Early automobiles were fuelled by mixtures containing benzene, and present-day petrol (gasoline) contains some benzene, although as a much smaller percentages than in some historical periods.

Apart from occupational exposure, benzene exposure may derive from combustion, both of domestic fuels and of tobacco.

The first observation that leukaemia risks may be inflated with benzene exposure was in 1928, in a case of acute lymphoblastic leukaemia. Corroborative evidence mounted until IARC classified benzene carcinogenic to man in 1982. Rinsky *et al* (1987), from a small study with relatively high exposures, established an exposure-response relationship between leukaemia and exposure to benzene, and this study has been subsequently refined, with both extended cohort follow-up and detailed retrospective exposure assessments. Plausible pathways and mechanisms for benzene's role of action have been proposed.

Rinsky *et al* (1987) studied the "Pliofilm" cohort, who were employees in a factory manufacturing an artificial barrier membrane. Others have studied the risks in, for example, shoe manufacture using benzene-based adhesives. These studies were typified by higher exposures, and there is now little dissent from the consensus that high exposures bring an increased risk of leukaemia. However, the question remained of whether leukaemia risks were elevated at lower exposures. For other lymphatic-haematopoietic cancers, such as non-Hodgkin lymphoma and multiple myeloma, there is little evidence for a relationship of risk with exposure even in those exposed at higher levels, so the question of risk at low exposures is generally understood to apply to the leukaemias only, and in particular to acute myeloid leukaemia.

Most of the studies that identified the risks in higher exposures were cohort studies, with mortality or registration of cancer diagnoses followed up over a period. Such studies often have rudimentary exposure assessment, so that internal exposure-based analyses are possible only at a crude level. The alternative, external analyses comparing cause-specific rates with reference rates can be difficult to interpret, because of the "healthy worker effect".

1.2 SELECTION OF STUDIES

Because of the benzene component in petrol (gasoline), workers involved in the distribution and marketing sectors of this industry, such as tanker drivers and loaders, forecourt attendants etc., were likely to have had exposures to benzene at lower levels than in some other industries. Petroleum industry workers were therefore worth studying, and there have been a number of cohort studies. Within a few of these, more detailed nested case-control studies have attempted to address the question of risks of benzene at a greater level of detail, focussing on leukaemia as an outcome, and, in some cases, other lymphatic-haematopoietic cancers. 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 these studies were broadly similar in showing little evidence of an effect of low benzene exposure, except for the Australian study, which reported a leukaemogenic effect at much lower levels than the other studies. All the studies used slightly different methods and procedures, and the reason(s) for the different results were not clear.

1.3 THE CALL FOR PROPOSALS

The Institute of Occupational Medicine (IOM) received a letter from the oil companies' European organisation for environment, health and safety, CONCAWE, inviting them to tender for a study to evaluate data consistency and quality in these four 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, since it was clear that a pooled study if practical would gain in power, to discuss aspects important for the possible pooling of the studies. The primary purpose was to elucidate the reasons for inconsistencies by characterising the similarities and differences among the four studies.

The call identified a number of factors that should be investigated and considered, including case definition and ascertainment, diagnosis of casehood and type, exposure assessment methods, selection and information biases, and potential confounding factors.

The call included explicit terms of reference for the work.

2 THIS STUDY

2.1 IOM/IRAS PROPOSAL

The IOM responded to the call with a written proposal and costing. The proposal was for a collaborative project, led by the IOM. Because the IOM's Dr John Cherrie was already advising the industry on this study, expertise in exposure assessment was supplied to the study group by IRAS at the University of Utrecht, who have particular expertise in occupational exposures, and with whom the IOM have collaborated on courses and studies over many years. Expertise in haematology was recruited by collaboration with a consultant from Glasgow University.

The proposal commented on the stated requirements, and on the relative balance between studying procedures and checking data-handling for errors. Our belief was then, and remains still, that the emphasis should be on the first, primarily because it would not be feasible to check every step of every data-handling process for every subject in the studies, but also in the expectation that serious systematic errors and endemic carelessness were both highly unlikely in such well-established investigative groups. This indeed had already been partially demonstrated by IOM staff in an earlier limited audit of certain aspects of the exposure assessment procedures of the UK study (Hurley *et al*, 1995). We therefore proposed that detailed data checking should be limited to those aspects where the work could be done centrally and in a cost-effective way.

2.2 OBJECTIVES

We proposed an audit that would concentrate principally on the aspects most likely to bear upon the principal questions, which here concern the comparability of results and the potential for pooling data sets.

We therefore suggested that the audit should focus on the following questions:

- Differences in definition of cases
- Patterns 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

Our proposal explicitly excluded the detailed checking of areas such as cohort enumeration, extraction of data from source, and tracing of vital status or cancer registration in national systems.

2.3 PLAN FOR VISITS

We proposed a strategic approach that was based on visits to a single central point for each study, to which the principal investigators (PIs) would bring all the necessary project materials for inspection. We recommended an initial fact-finding visit by the project leader to each site, to set up liaison, and to make adequate preparation for the main visit. The second visit, lasting up to five days per site, would be carried out to inspect and assess the available records and documentation on detail, and would be made by two of the team, supplying expertise in both epidemiology and exposure assessment.

We highlighted that important preparation for these visits would include the collation of existing documentation including research reports, published papers, and any existing review reports.

We also envisaged that, to assist in maintaining consistency between visits to the different sites and the studies associated with them, we needed to standardise the topics addressed. To achieve this, we drafted and included in our proposal a structured list of topics and specific questions; and proposed that a detailed protocol for the visits should be developed and agreed in advance with the PIs and CONCAWE's scientific advisers.

We identified the need to answer the questions quantitatively, e.g. number or proportion of non-responders, number of subjects with missing data, etc. We also envisaged investigating some 5-10% of suitably chosen cases and controls in depth, checking data against sources and the results of calculations for accuracy.

We noted that there would be issues of confidentiality in handling personal data with identifiers, and agreed in principle that our purposes could be met with data anonymised to a study id number.

2.4 THE CONTRACT

Our proposal was accepted, and a contract entered into for the work. In the process, CONCAWE asked us to revise the proposed methods, and in particular to remove the initial site visits. This was particularly because of the large costs involved in visiting geographically distant sites. We agreed that an alternative approach, based on bringing together all the PIs to meet with the study team in a central location, would still allow the team to meet the PIs and to discuss and develop the protocol for the visits, although it would necessarily preclude any inspection of the site record-keeping facilities.

2.5 PREPARATORY MEETING

As part of the preparation for the study, CONCAWE contacted the PIs and invited them to participate in the review. Agreement was received from the PIs of the Canadian, UK and Australian studies. The former PI of the US study replied that he was no longer employed in that capacity, and he indicated that did not know where the records from that study were held. As a result, it was decided that there would be no site visit or review of records from that study, but that a review of the study would be made simply on the basis of available technical reports and published papers. Because the US study had not made detailed estimates of individual exposures to benzene, its loss was felt to be less damaging than would have been the case for any of the other three studies.

It was also noted that the Australian study was controlled from two separate locations, with the Health Watch cohort study now run from the University of Adelaide, and the case-control study run from Monash University, Melbourne, so it was agreed that it would be necessary to visit both sites during the Australian visit.

2.6 VISITS

The site visits were carried out by Brian Miller, IOM and Wouter Fransman, IRAS, over the summer of 2004: They visited EMBSI at the EXXON offices in Clinton, NJ 27-30 July, University of Birmingham 24-27 August, and Monash University, Melbourne 22-27 September. Brian Miller visited University of Adelaide 5-7 October.

At all visits, the PIs and associate staff made all possible efforts to accommodate the reviewers, and we are most grateful for their cooperation and help.

2.7 DOCUMENTATION

Prior to the visits, we obtained either from the PIs, or through CONCAWE or their advisers, or through our own library facilities and information resources, as full a set of background documents as we could (see References, chapter 7). These included:

- papers from peer-review journals;
- technical reports;
- procedure manuals.

We supplemented this during the visits with additional examples and more detailed documentation, where these were available only locally.

We also obtained, as planned, material for the cases and controls selected as a sample from each study, fully anonymised. As described separately for each study, there were instances when material was not available on site and could not be supplied. We discuss the implications of this below.

3 DATA QUALITY AND RELIABILITY

3.1 NOMENCLATURE

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

- IOL: study on Canadian employees of Imperial Oil and associated companies (PI R Schnatter)
- IP: UK study of workers from four companies affiliated to the Institute of Petroleum (PI T Sorahan, formerly L Rushton)
- API: US study of petroleum industry worker (PI O Wong)
- HW: Australian Health Watch study (PIs R Gun – cohort study, D Glass – case-control study)

3.2 COHORT DEFINITIONS

The IOL cohort was around 34,000 subjects, of which 6672 were identified as belonging to the petroleum distribution and marketing sectors, at 226 sites. Inclusion required one year's employment before 1983. All selection was based on company employment records, and there was no objective external data source against which cohort completeness could be verified. However, the investigators rightly note that employment record systems, which form the basis of payment structures, tend to be well maintained. Where sample checks were carried out between computerised and paper sources, they showed almost complete agreement, and the PI's team have no concerns regarding cohort completeness. Because the original records from which the cohort was selected were not on site, and because the transfer was electronic, it was not possible to make any direct verification of this claim, but the review team has no reason to doubt it.

The IP cohort was defined around UK petroleum centres within three major companies, later joined by a fourth. This covered the great majority of the industry in the UK. Eligibility required one year's service before the end of 1975. Cohort identification was based on company employment and pensions records, which may be expected to be accurate and complete. We were able to inspect the large number of boxes containing the standard data entry forms, but not, of course, to check these with the original sources. We therefore accept the study team's views that the cohort is essentially complete.

The API cohort included land-based and marine workers, with the case-control study nested within the land-based portion. Recruitment was again based on company records, and for the same reasons as for the other studies these are expected to have been reliable. However, eligibility depended on a judgment that the workers had opportunity for gasoline exposure, made locally, and the investigators noted that this might have led to the loss of a small proportion of workers in jobs with low exposures at some sites.

The HW cohort was intended to include all industrial workers from participating companies, which constituted almost the entire Australian petroleum industry. Only city centre offices and a few very small sites (employing fewer than 10 persons) were excluded. The cohort was unique among the present studies in that it required the voluntary participation of its recruits, because data collection included questionnaires to the cohort members. Some have suggested that this voluntary aspect may have created a bias in the study composition, but his criticism appears to be based on a misunderstanding of the recruitment process. If as suggested a subject could join the cohort when and if he pleased, then indeed it would be difficult to judge the completeness of the cohort. However, the fact is that both the initial phase of the recruitment and the subsequent periodic additions to the cohort were based on company employment records; and the PI reports that co-operation by subjects requested to participate was, in the early days when most of the subjects were recruited, well in excess of 90%. While it is not possible to verify, it seems unlikely that this small loss from the target population through non-response could have been so systematic as to seriously skew the results of even the cohort study, and doubly unlikely that it could have had an effect on comparisons between identified cases and age-matched controls. We therefore see no reason to be concerned about cohort incompleteness in the HW study.

Another suggestion that has been made about the HW study is that an active blood screening programme for benzene-exposed workers might have preferentially identified cases with benzene exposure, leading to a selection bias. However, on checking by the study team it was found that no case was identified through the screening programme, so no bias can have arisen from this source.

3.3 CASEHOOD DEFINITIONS

Cases in the IOL study were identified from the cohort study data files containing the results of the tracing exercises. Causes of interest, identified by their standard codes in the International Classification of Diseases system, revision 8 (ICD8), were

- leukaemia (ICD8 codes 204-207)
- multiple myeloma (ICD8 203)
- non-Hodgkin lymphoma (ICD8 200, 202.0, 202.1, 202.2, 202.9)

Restriction to subjects who had ever worked in the marketing, distribution, pipeline or marine segments yielded, between 1964 and 1983, 16 leukaemias, seven multiple myelomas and eight non-Hodgkin lymphomas. It was not possible at the time to obtain cell typing for the leukaemias. With such a small total number, analyses of subtypes would anyway have had very low power.

The IOL study team tracked the performance of the vital status ascertainment through the Statistics Canada (SC) systems, on 757 deaths already known at the start of the study (Schnatter *et al*, 1990). SC systems correctly detected over 93% of these, including 97.6% of those deaths that occurred in Canada. Specificity was almost 100%. It is therefore reasonable to conclude that non-ascertainment of cases is not a cause for concern in the IOL study.

Cases in the IP study were defined as deaths or diagnosed cases of leukaemia only (ICD9 204-208), traced through the UK's death registration or cancer registration systems via the National Health Service Central Register. This is the standard method for tracing vital status in UK cohorts, with a tracing rate typically 95-97% for deaths. Thus non-ascertainment of mortality cases was not a cause for concern. Because complete cancer registration built up from its introduction in 1971, it is possible that some non-fatal cancers from the early 1970s may have been missed. However, it seems highly unlikely that could be a source of bias for the case-control relationships.

A total of 93 leukaemia cases were identified in the IP study, and cell types were distinguished in the analyses. Distinctions were made on the basis of the descriptions on the death certificates or cancer registrations. In the 40 cases found in both systems, the death certificates usually gave the more detailed description. Confirmation of cell type from histo-pathology departments was achieved for 38 cases prior to the publication of the results. In no case did a histology report result in a change of diagnosis, so it is plausible that the diagnoses for the unconfirmed cases are similarly reliable.

Rushton and Romaniuk (1997) give a detailed breakdown of the cell types involved. In some cases, it seems that the information available was not sufficient to make a distinction between cell types (five cases), or between acute and chronic (four additional cases); these indeterminacies are the result of an inability to trace clinical details. Analyses were carried out separately for different celltypes. The authors note (p162) that one case of acute myeloid leukaemia (AML) also had chronic myeloid leukaemia (CML) on the death certificate but was classified as acute myeloid leukaemia for this study; we would query this decision, and would usually treat such a case as inherently chronic. In addition, the five cases classified as "other " for cell type included one case of acute erythroid leukaemia, and it is arguable that this case could have been included as an additional AML; indeed, the authors note that three acute cases might have been included if the category had been defined as "acute non-lymphocytic" rather than "acute myeloid".

The API study traced vital status for their cohort through US systems, beginning with the Social Security Administration, the US Death Master File, and the National Death Index. Copies of death certificates were sought from state agencies. The authors do not discuss the likely rates of non-ascertainment in these systems.

The API case-control study includes cases series for kidney cancer, leukaemia (with sub-analyses for AML), and multiple myeloma.

The HW case-control study defined the outcome as a "newly diagnosed lympho-haematopoietic cancer". Because of the participatory element of the study, and to satisfy stringent Australian ethical requirements on data protection and confidentiality, the case had to be reported to Health Watch either by himself or his family, otherwise the researchers were precluded from verifying case status, unless the case was already dead or lost to follow-up in the cohort study. Some protection against non-ascertainment was given by the high participation rate, and in the event only one otherwise eligible case traced through the cancer registry had to be omitted because he had not self-reported. Therefore non-ascertainment is not a cause for concern in the HW study.

It has been suggested that the HW study may have suffered from the fact that cancer registries in Australia were in development during the early stages of the HW study. At worst, this could have led to an under-identification of cases, but this is unlikely to have been extensive, since this was in a period when prognosis for leukaemias was much worse than in present times, and death registration would have been near complete. While loss of a few cases might have reduced the power of the case-control study, there is no reason to believe that the probability of loss would have been related to exposure. Even if some early cases were missed, we therefore do not believe that this could have distorted exposure-response relationships based on comparisons within the case-control sets identified and analysed.

HW identified 79 cases, of which 33 were leukaemias. Medical information for all cases where diagnosis was uncertain was reviewed by an expert haematologist, confirming or revising the cell type involved.

3.4 CONTROL SELECTION

For the IOL study, male controls were selected from the cohort database at a ratio of four to one, matched by decade of birth and alive on or after the cases' dates of death. During the visit, we were told that matching was not individual, but category matched within decade, but that the analysis used software for individually matched data. This is somewhat unusual, in that the textbooks advise constructing analyses that reflect the matching criteria exactly. However, it is accepted that the distinctions between individual and group (stratified) matching are frequently blurred in case-control studies. The end result is a slight approximation that is unlikely to have affected the results noticeably, compared (say) to analysing matched data as if they were not matched at all.

In the IP study, controls were selected without replacement at ratio of 4:1, from the same company, and with a year of birth within three years of the case's. They had to be alive at the time of the cases's occurrence. Data were analysed as matched. The IP investigators have noted that their checking identified retrospectively a small error in the matching at one site. The PIs investigated, but it is not clear how this arose. The invalidly selected controls were omitted from the analyses, along with one case that then had no valid controls. This will have resulted in a very small loss of power, but is highly unlikely to have introduced any detectable bias.

In the API study, controls were selected from the cohort study database, at a target ratio of 5:1, matching on company, sex and with a date of birth within two years of the case; and they must have been alive at the time of death of the corresponding case. Analyses were done making appropriate allowance for the matching.

In the HW study, controls were selected from the database of the cohort study at a ratio of 5:1, matched on sex and year of birth, and sampled with replacement. Almost all the cases and their controls were selected while the study was held at Melbourne University, and only three new leukaemia cases were added after the cohort study was transferred to Adelaide. We saw the database macro employed during the Adelaide visit. Analyses were done making appropriate allowance for the matching.

The API and IP studies explicitly matched on company, whereas the HW did not. The IOL study was within a single company, so the subjects were automatically matched on company. Quibbles might be raised about what "company " means here, since any large company may be an amalgam of previous acquisitions, and may have many trading subsidiaries that are set up as separate companies in law. In the present case, all the controls came from the same broad industry, but exposures were more likely to vary by industry sector within company than between companies. It is unlikely that matching within company would have constrained the available range of control exposures in such a way as to distort the exposure-response relationships, and we do not expect that the decision whether to match on company will have made any important difference to the results.

Apart from the points noted above, the control selection was standard. In all we saw no evidence that differences in the selection of the controls were likely to have had any serious effect on the case-control study results.

3.5 WORK HISTORIES

In the IOL study, work history data were extracted from the company personnel records by IOL staff, working blind as to case-control status. Data extracted was for events defining changes in employment or exposure status, such as hire, transfer, promotion, loan and termination. Recorded events with no impact on exposure (e.g. organisational, changes in pay scales) were not extracted. Because this process took place in Canada, we were not able to inspect the source documents. However, there was a detailed written protocol, and pre-prepared forms for data collection, and we have no cause for concern regarding the procedures used. Some work history data came from an electronic system that had been introduced from the 1960s, but the format of the data was identical with the hard copy systems, and the distinction is not important.

Initially, job descriptions were in text form, and not coded. Job titles were subsequently clustered into a set of job title categories. Where possible, data were checked with medical records, but in general these proved to be less detailed that the personnel data.

These procedures, as described, would appear appropriate for the data sources, and give us no cause for concern.

In the IP study, the cohort data on occupation had start and finish dates, but only the last job was identified. The required detailed data on work history differed in its availability across the four companies included, and what was retained was variously stored in hard copy and/or microfiche or microfilm. Staff extracting the data worked blind to case/control status. We did not have access to the original sources, and did not attempt to read any of the microfiche or microfilm. In addition, since the IOM's original audit, the original documents that had been extracted, copied or printed were no longer present in the filing cabinets, although they had been seen by Hurley and Cherrie during the IOM's earlier audit. All that remained was material for the small number of subjects from Company C. L Rushton suggested from memory that there had been concern for the confidentiality of the material, since a proportion had originated from medical records; and that the source documents may have been returned to the company medical officers. While this is plausible, we saw no documentation of this process, and cannot tell exactly when it took place, although it was presumably between the IOM audit and the transfer to Birmingham. We note that the detailed checking during the Hurley & Cherrie audit did not find any serious problems with the quality of the work history data.

Some imputation was employed where work histories were incomplete, but codes were used to identify incomplete records, allowing analyses including and excluding these records.

In the API study, the reports record that all the work history data came from the original cohort data, which had been extracted from employment records. The descriptions of missing data are sketchy, but suggest that this was not seen as a large problem. Without either interviews with study staff or access to source documents, we have not been able to verify this impression.

In the HW study, work histories were taken from the data already held in the cohort database, taken from self-reports from the periodic surveys, and cross-validated with company records. We were therefore not able to check the data with the source documents, although we have, for our representative sample, copies of the completed detailed questionnaires. There was some missing data in the final set, but as described in the detailed report of the site visit the proportion of incomplete work histories is now estimated as around 2%, so its impact is likely to be negligible.

3.6 DATA ON POTENTIAL CONFOUNDERS

In the IOL study, other data were collected, where available, from company medical records; and 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. Availability was very patchy. In particular, smoking data were available for only a small minority, and were unknown for almost 90% of cases and controls.

No data on smoking or lifestyle habits were available in the API study.

For the HW study, the cohort database included data from the health survey questionnaires, on smoking habits, typical alcohol consumption and on previous employments. The smoking data were relatively detailed, including amounts smoked.

3.7 AVAILABLE MEASUREMENT DATA ON BENZENE CONCENTRATIONS

The exposure assessment methodologies that were used in the IOL, IP and HW case-control studies were all based on an approach that was designed for the IOL study as described by Armstrong et al. (1996). This methodology for retrospective exposure assessment starts from "base estimates". This base estimate is an arithmetic mean exposure based on measurement series, assigned to typical jobs and periods. These base estimates were subsequently adjusted for modifying factors (K-factors) assigned by experts, to allow for differences in local conditions and practices.

3.7.1 IOL study

The IOL cohort consisted of a population from petrol distribution sites. A large proportion of subjects in the case-control study had a work history of only background exposure (≈41%), while only 15% of the population had a work history composed entirely of exposed jobs in petrol distribution with exposure to benzene. The remaining 44% had work histories with a combination of periods of exposed jobs and unexposed jobs to which a population background level was assigned. The combination of relatively low benzene exposures and small numbers of leukaemia cases was likely to make it more difficult to detect a dose-response relationship.

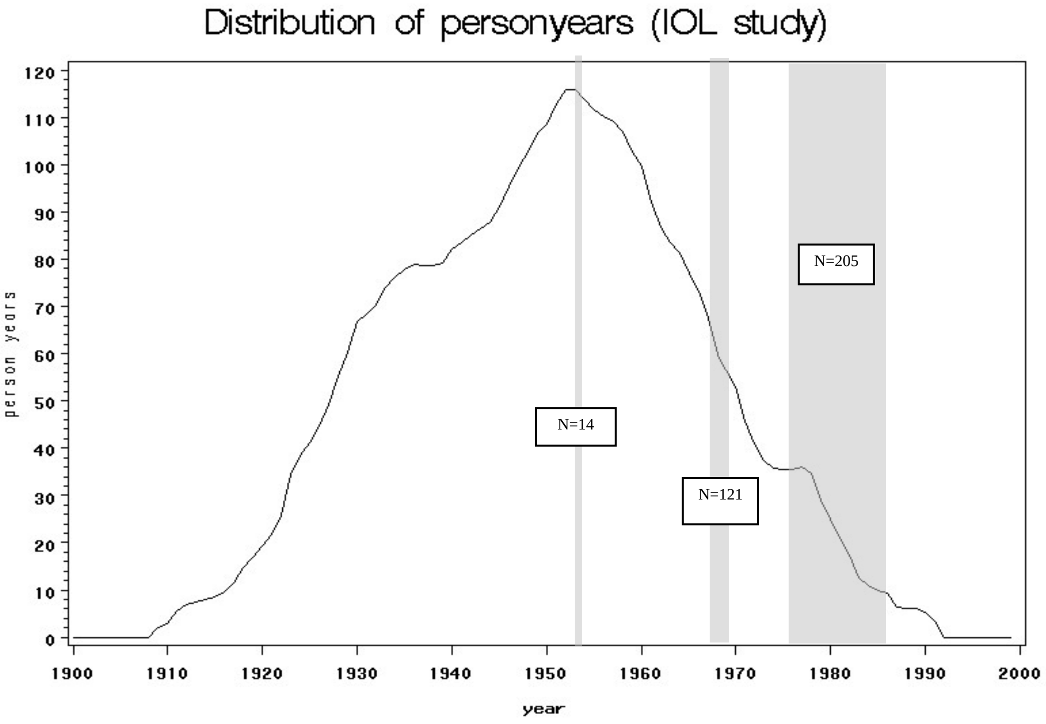

Figure 3.1
Overlap between available measurement data (shaded) and the distribution of person years from 1900 to 1999 for the IOL case-control study (N=number of measurements available for BE derivation)

Exposure data came from IOL-company industrial hygiene survey reports, complemented by published papers and CONCAWE reports. These reports describe exposure data that originated from the period between 1975 and 1986. Some historical data were used for 'loaders' (company report, 1953) and 'air&service station attendants' (Parkinson, 1971). Fifty percent (seven out of fourteen) of base estimates were constructed using less than 10 measurement data points, and some base estimates were based on only 3 or 4 data points. Base estimates were based on both full-shift and short-term measurements and some base estimates were converted from total hydrocarbon (THC) results, as is described in greater detail in section A1.7.4 and is presented in Table A1.1 (Appendix A1). It is not made explicit how these different types of measurements were used to calculate base estimates.

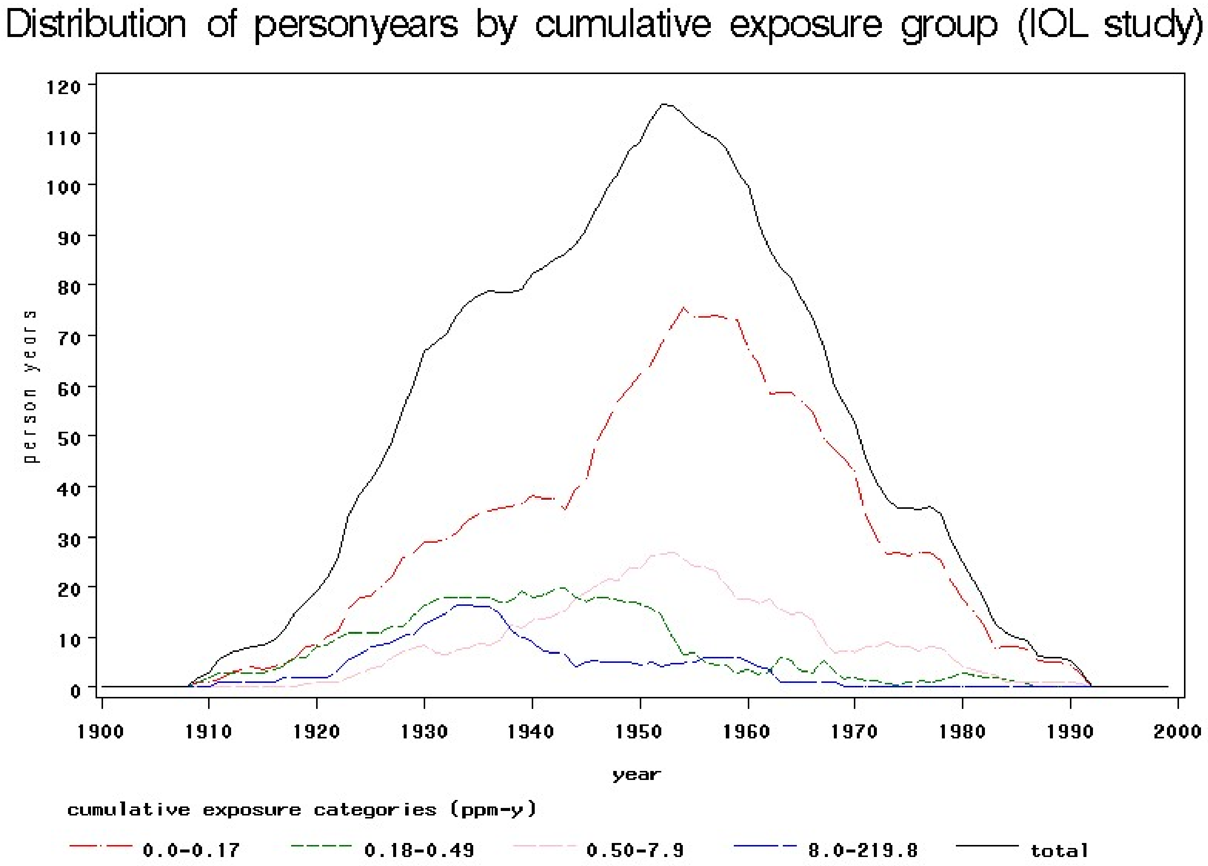

Figure 3.2
Distribution of person years from 1900 to 1999 for the IOL case-control study by cumulative exposure category.

The cases from this cohort included in the case control study emerged between 1964 and 1983. So, the exposure for these cases occurred between 1944 and the late 1970s, assuming a latency of between 5 and 20 of exposure. Although the distribution of cases over time was not evaluated, this requires exposure estimation over a broad period and extrapolation to early exposure periods. This is made clear in greater detail in Figure 3.1, in which the distribution of person years from 1900 to 1999 is presented. The shaded areas in this figure reflect the time periods for which measurement data were available (N is the number of measurements used for BE derivation), which resulted in a 9.9% overlap between person years and exposure measurement data. This illustrates that current exposure levels had to be extrapolated back over a considerable period (several decades) in order to be able to assign quantitative exposure estimates to the majority of person years in this study.

In Figure 3.2 the distribution of person years by cumulative benzene exposure category (by quartiles) is presented. This figure shows that the subjects with highest cumulative exposure are mainly represented in the early periods (1920-1940) of the study, for which the industrial hygienists were least certain about their exposure estimates (Table 3.1). Table 3.1 also shows that the industrial hygienists are most certain about the low cumulative exposure levels. These low levels mainly involved the various categories of workers with ‘population background’ and ‘site background’ levels. The hygienists were least certain about the higher cumulative exposure levels where most extrapolation was needed beyond the periods for which measurements were available.

Table 3.1
Certainty scores assigned to exposure estimates by industrial hygienists for each cumulative exposure category in the IOL study.

	Certainty 1 “best guess”		Certainty 2 “substantial info”		Certainty 3 “most certain”	
cumulative exposure (ppm.years)	**N**	**row%**	**N**	**row%**	**N**	**row%**
0.0-0.17	119	15.5%	250	32.4%	401	52.1%
0.18-0.49	36	43.4%	29	34.9%	18	21.7%
0.50-7.9	70	40.0%	101	57.7%	4	2.3%
8.0-219.8	22	81.5%	5	18.5%	0	0.0%
	ppm.yrs	**row%**	**ppm.yrs**	**row%**	**ppm.yrs**	**row%**
Total cumulative exposure (ppm.years)	1184.7	74.4%	382.3	24.0%	25.6	1.6%

3.7.2 IP study

In the IP study most exposure measurements that were used came from cohort company industrial hygiene reports. A few base estimates were derived from CONCAWE reports and for the job group ‘terminal operator’ additional monitoring was carried out for the case-control study. Most of those exposure reports were not available in Birmingham during the site visit and therefore we cannot say much about the quality of the data that were used. However, during the IOM audit (Hurley *et al*, 1995) carried out in 1995 in Nottingham, these exposure measurement reports were all available in Nottingham as well as the computer spreadsheet files with the exposure measurement results that were used for base estimate derivation. At that time the auditors concluded that the derivation of base estimates had been carried out in a careful manner and no data entry errors were found. However, they recommended that all reports collected as potential data sources for the IP project should be catalogued and copies should be archived (IOM audit: Hurley *et al.*, 1995), of which no evidence was found in Birmingham in 2004. A possible reason for the missing industrial hygiene survey reports in Birmingham is that company reports were returned to the respective companies at completion of the study. The cases that were included in the IP case control study occurred over a somewhat longer period between 1950 and 1994. The potential exposure extrapolation is therefore somewhat more extensive compared to the IOL study. A similar analysis of person years distribution versus exposure data could be made for this cohort. From the limited SAS datasets that were available for the IP study, we were able to create a similar graph for the IP study, in which the overlap between exposure measurements and the distribution of person years is presented, which appeared to be 9.1% (Figure 3.3). From the list of over 200 company industrial hygiene reports that were considered for use in the exposure assessment we do know that all these company industrial hygiene reports were from the period between 1969 and 1993 (although we do not know which reports were used). If complementary literature data were used, we assume that these would be from the same period (or later). Therefore, in figure 3.3 we shaded the time period from which we can be almost certain that measurement data were used for base estimate derivation. This illustrates that extrapolation was needed for the majority of person years in the IP study.

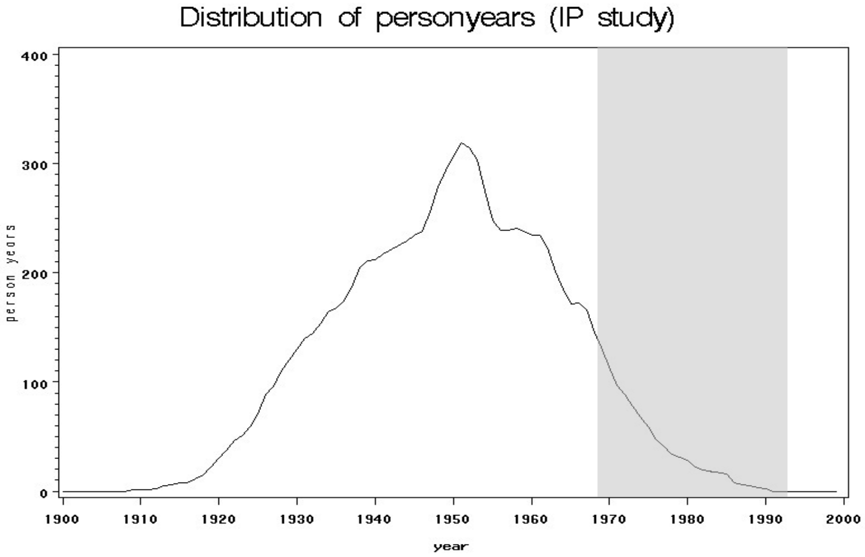

Figure 3.3
Distribution of person years from 1900 to 1999 for the IP case-control study and time periods of available measurement data (shaded).

3.7.3 HW study

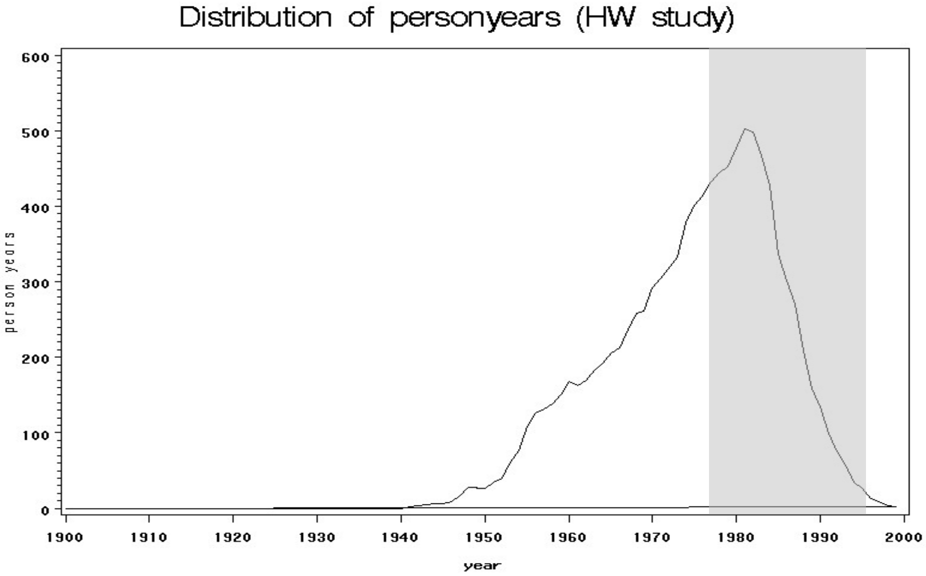

Figure 3.4
Overlap between available measurement data (shaded) and the distribution of person years from 1900 to 1999 for the HW case-control study

The HW cohort consists of mostly occupationally exposed workers, with a relatively small reference population, which leads to a relatively high power to detect associations between exposure and disease. The cohort was enrolled in a different period and cases were diagnosed in more recent years between 1981 and 2000.

In the HW study most exposure data were derived from company data and were also supplemented by CONCAWE reports and published papers. Most data were collected in the period between 1978 and 1995. The main difference from the other studies was that the HW study used task-based exposure data, which were used to estimate benzene exposure levels for a task, which were summed over a job after adjusting for several modifying factors (K-factors). For some task estimates only few measurement data points were used to construct the task estimate. Figure 3.4 presents the distribution of person years for the HW study and the time period with available measurement data for base estimate derivation. This overlap appeared to be 45.5%, which demonstrates that measurement data were available for a relatively large proportion of person years in this study, so that less extrapolation was needed to assess retrospective exposure levels in the HW study relative to the two other studies.

3.7.4 API study

In the API study, exposure measurements from the four participating companies collected between 1975-1985 were used to quantify exposure intensities associated with a specific combination of job tasks, types of worksites, and types of products. Because we did not visit this site and the published paper and report on the nested case-control study do not reference any of the exposure measurement results, we cannot comment on the exposure data that were used in this study.

3.8 EXPOSURE ASSESSMENT & ASSIGNMENT

The exposure assessment methodologies that were used in the IOL, IP and HW case-control studies were all based on the method that was designed for the IOL study as described by Armstrong et al. (1996). This methodology for retrospective exposure assessment is based on a “base estimate”, derived from arithmetic mean exposure measurement results, which was assigned to each line of work history (for combinations of typical jobs and time periods), and subsequently adjusted for several modifying factors (K-factors) assigned by expert knowledge to adjust for deviations from the base estimate situation (such as: loading technology, temperature, fuel volatility, benzene content, product mix, work practices, task frequency&duration, number of loads per day, ventilation). This resulted in a workplace exposure estimate (WE) for each line of work history: $WE = BE * K_1 * K_2 * K_3 * ... * K_n$. The workplace exposure estimates multiplied by the time spent in each of the recorded jobs and subsequently summed for each individual resulted in the cumulative exposure (CE) in ppm-years.

Throughout the three case-control studies an evolution of the exposure assessment and assignment methodology over time is apparent, which is mainly evident in the structure and detail of the exposure data and K-factors that were used. The approach of base estimates is similar throughout the three studies (IOL, IP, and HW), although the amount of data, on which the estimates are based, differs as is described in section 3.7.

Also the use of the exposure modifiers is different. In the IOL study, the way exposure modifiers were used to obtain alternative exposure estimates has not been made explicit and the assignment of exposure modifiers (K-factors) has been performed by an expert panel of industrial hygienists. Some of these modifiers have a physical basis (effect of temperature, fuel volatility), but have not been validated in this specific context in none of the three case-control studies. A limited validation exercise was performed in the IOL study with 35 constructed work history lines with measured benzene exposure levels. This validation exercise resulted in an average percent difference between estimated and measured exposure of 23% (range 0%-130% difference). Moreover, it is surprising that no attempt has been made to obtain estimates for some of the K-factors on the basis of a multivariate analysis of the available exposure data and contextual information. In the IP study the exposure assignment process is more structured in that information about historical process changes (from interviews, etc.) were stored by company and by site, so that modifying factors per work history line could be sought in this information. However, generally speaking, the methodological approach was similar to the IOL study, as was the assignment of K-factors. The HW study is similar to the IP study in that historical information is stored by company and site. In addition, the HW study has obtained detailed information about the set of tasks that was performed within a certain job title in a certain company in a certain time period, how these tasks were performed (technology) and for what percentage of their time workers performed those tasks. Therefore, there was an extra level of detail added to the exposure assessment process, in which an extra task estimate (TE) was calculated (base estimate multiplied by K-factors) for each task that were subsequently summed for a job based on the time spent on individual tasks. This process is described in greater detail in section A3.7.3 (Appendix A3). Because back-extrapolation was relatively limited in the HW study, the use of K-factors was also more limited compared to the two other studies. The task-based approach of the HW study and the use of arithmetic mean (AM) instead of geometric mean (GM) for the task estimates could have led to biased estimates and overestimation of shift-long average exposure levels as for instance described by Seixas et al. (2003) and Coker (1984).

From the available literature, the API study seems to use a similar method to calculate cumulative exposure as in the HW study, in that it uses task exposure levels, which are time weighted for a certain job (the so called task-TWA model), although only for total hydrocarbon exposure. However, the API study does not seem to use a similar method to adjust for deviations from the exposure estimate (modifying factors) like the other studies, but is more focussed on the time period in the work history and the loading technology used in that specific time period.

3.9 STATISTICAL ANALYSIS

3.9.1 Overview

All the studies used techniques appropriate for matched case-control studies, and we have no serious concerns with this approach. However, there are a number of factors that affect the outcomes in case-control analyses, and a number of criticisms have been made by others, particularly of the HW study results.

3.9.2 On the statistical analysis of case-control studies

The statistical analysis of case-control studies is based on comparing the attributes of cases with those of controls; and where matching is used, the comparisons are made within each set defined by the matching. It is well known that that the analysis of matched data needs to respect the matching, and that if it is ignored in the analysis the estimates of relative risk may be biased.

Where data are individually matched, the correct analysis depends on the idea of conditional probability, and may be exemplified as follows; given a matched set of case(s) and control(s), and given that they are matched on some variables that might determine risk, is there any imbalance between their values of other, non-matching variables that might explain why a case is a case and a control is not? Quantitative evidence of any difference is then amassed over all case-control sets to provide a pooled estimate of relative risk. Variants of this approach can be used to account for group matching within strata.

It is also well known that case-control studies cannot, in themselves, provide information on absolute risk, only on relative risk. When as here we are interested in relating relative risk to levels of a quantitative exposure variable, difficulties arise as to how to express that relativity and how it varies with exposure. One way is to calculate a regression coefficient (usually within a conditional logistic model) that quantifies how the relative risk (on the log-odds scale) increases per unit increase of the exposure. This approach is open to the criticism that the usual model assumes a linear relationship on the log-odds scale.

Another approach is to group the exposures of the cases and controls into categories, and carry out an analysis that estimates how the log-odds differ between categories. Since we have no absolute measure of risk, we can only compare the relativities between categories, and we have to select one of the groups as a baseline, assigned an arbitrary level of unity. Usually we choose the lowest level, and then estimate the risks in the other categories relative to this baseline. Problems with this approach are twofold. The first concerns the size of the baseline group. If it is too small, it may be highly susceptible to sampling variation, and may be poorly determined; if too large, it may be inhomogeneous, and cover a range of relative risks, so that it is not a true baseline.

The second objection to the grouped-exposure approach is that the interpretation of the relative risk coefficients is usually made by comparing the coefficient to its standard error. This implies a comparison of each level with the baseline, and these comparisons cannot be statistically independent because they all use the same baseline. As far as we can see, none of the discussions of statistical significance deal with this point, or with the need to adjust the significance level for the multiple comparisons usually made.

Recent advances in statistical regression methods include the idea of Generalized Additive Models (Hastie and Tibshirani, 1990). These extend standard regression models by including non-linear spline curves. This would enable a plot to be made of how relative risk varies with exposure, with the form of the curve being dictated by patterns in the data rather than some pre-chosen functional form. Application of this technology to conditional logistic models may be technically difficult, but it is likely to be possible (we know it has already been used for Cox proportional hazards models), and will be worth investigating if a pooled analysis is undertaken.

3.9.3 A suggestion for graphical display

One of the biggest problems with the interpretation of the matched case-control data in these studies is that the complex structure of the data makes it hard to get a picture of what they tell us; indeed, there are few graphical presentations of the data in any of the reports that address relative risk.

In attempting to understand what is going on in the Australian study data, we have tried to design a plot that would give some at least some insight. We start from the knowledge that, under a null hypothesis of no relationship between exposure and risk, the exposures of cases and controls should not differ except at random. In the simplest case, of 1:1 matching, we could graph each case exposure against that of their matched control on a graph with equal scales. In the null case, points would be equally distributed around the line of equality. If the matching variable were partially correlated with exposure, as we would expect age to be, the points might cluster in a band around the equality line, but the points should still distribute symmetrically around that line.

We can apply this idea to the HW study data, although here we note that each case has five corresponding controls. Figure 3.5 shows a plot of the cumulative benzene exposures of cases against their corresponding controls, on logarithmic scales; the points are indicated by crosses. The 45° line of equality is shown, as are grid lines showing the boundaries of the logarithmically increasing exposure categories used in the paper. The pattern is diffuse and complex, although there is a clear impression of more points to the left than to the right of the equality line. Superimposed on the plot and indicated by black circles are the arithmetic means of the exposures of each set of controls against their matched case.

This plot of cases against control means is shown in isolation in Figure 3.6. We can count that there are 21 points to the left of the equality line, and 12 to the right, suggesting that indeed the cases have higher exposures than the controls. In addition, taking the perpendicular distance of the point from the line as a measure of how different the exposures are, the eye gives the impression of greater distances at the higher exposure levels.

This is seen more clearly if we rotate the picture clockwise by 45º, to give Figure 3.7. The horizontal distances now represent, on the log scale, the ratio of the case exposure to the mean of the controls' exposures. The y-axis is, on the log scale, the geometric mean of the case and control mean exposures. The graph also shows a 7-point running average of the ratios. This emphasises that there is little evidence of an exposure-response relationship at the lower levels of exposure, but that the evidence of an effect in the higher exposures is consistent. As an alternative summary, Figure 3.8 shows (in a different orientation - rotated by 45° and reflected in the line of equal ratio) the result of fitting a generalized additive model to the log-log relationship between the ratio and the geometric mean, with a smoothing spline to define the curve. Again, this shows that all the evidence for an effect is amongst those case-control sets with the higher exposures.

We note that these plots do not reflect exactly the relative weights of the different case-control sets in the formal analyses, and that the smoothed summaries in Figures 3.7 and 3.8 will not reflect exactly the shape of the exposure-response relationship that might emerge from such a formal analysis. It is also true that the graphs differ if the controls are summarised by a geometric rather than an arithmetic mean; Figure 3.9 shows that this reduces the impression of an effect only at the higher exposures. Figure 3.10, which is based on linear rather than log scales, and thus uses differences rather than ratios for comparison, emphasises a pattern of increasing differences in exposure (which is consistent with a constant ratio). However, all these graphs are consistent in displaying that what effect there is is concentrated in the case-control sets with the highest exposures. We expect that this type of graphical display may be useful generally, as a method of examining the consistency of case-control differences through the range of an exposure variable, at least in situations where there are no additional covariates with strong effects.

Incidentally, Figure 3.6 shows very clearly that the three cases whose exposures fell in the first exposure category (<1 ppm.year) all had exposures lower than the mean of their matched controls, and in one case considerably lower. This is consistent with the suggestion that the first exposure group represents, possibly by chance, an artificially low baseline; if so, the relative risks in all the other, higher exposure groups, which are expressed in comparison to the baseline, will all be inflated.

We have now been able to apply the same graphical approach to the summary data from the IP study. Figure 3.11 shows all the data, with the exposures of the cases on the y-axis, and the individual controls and their arithmetic means plotted against the x-axis. Figure 3.12 shows the cases plotted against the control arithmetic means alone.

Figure 3.13 shows the same data points as Figure 3.12, again reoriented. Points where the exposure of the case exceeds the arithmetic mean of the controls plot to the right of the vertical line, and the distance from the line is proportional to the log of the ratio of case to control. The dashed line is a 7-point moving average of those distances.

Figure 3.13 suggests that there is no evidence for any excess risk of leukaemia at lower exposures, and that the only evidence for an increased risk lies in a few case sets with higher exposures, at the top right of the figure.

Figure 3.14 is similar, but plots the cases against the geometric means of the controls. As before, the dashed line for the moving average shows some movement to the right compared to Figure 3.13, but the suggestion is still that there is no evidence for increased risk at the lower exposures. We did not have access to the corresponding individual data for the IOL study, and so were not able to construct comparable graphs for that study.

We believe that any re-analyses of the data sets or pooled analyses should make extensive use of graphical presentations to complement formal model-fitting analyses. We present the graphs in Figures 3.5 to 3.14 as a first attempt to look at the individual data within case-control sets; as one example of what might be done in this direction. It is probable that other designs could give different (and better) insights, and there is scope for ingenuity in creating new designs or adapting old ones.

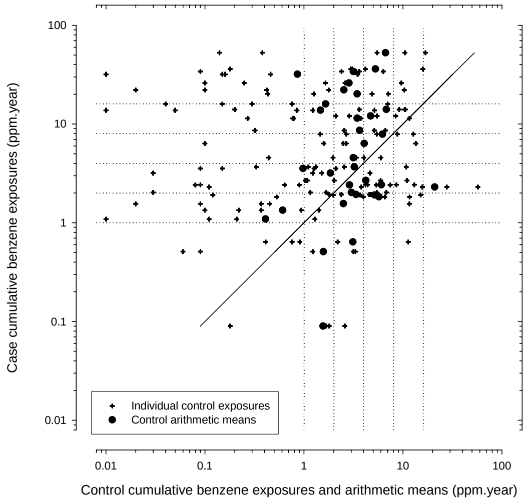

Figure 3.5
HW data: Cumulative benzene exposures of cases against control individual exposures and control-set arithmetic means

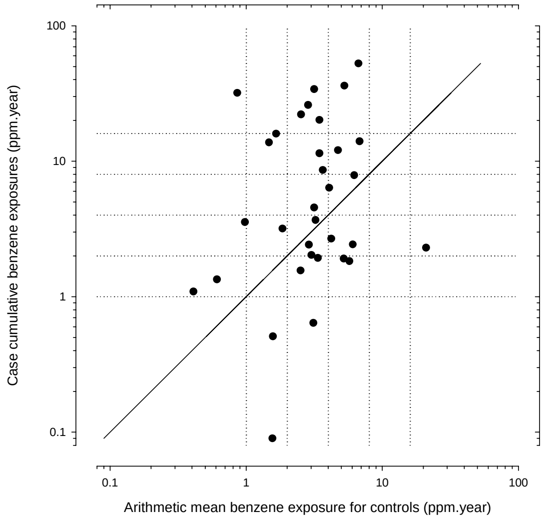

Figure 3.6
HW data: Cumulative benzene exposures of cases against control-set arithmetic means

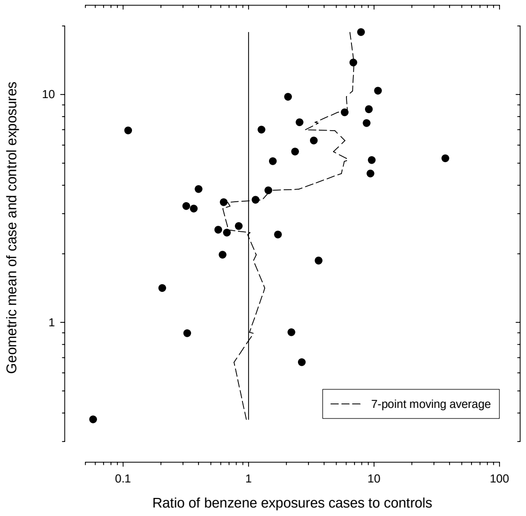

Figure 3.7
HW data: Ratio of case to control-set arithmetic mean exposures against geometric mean of case and control-set arithmetic mean exposures

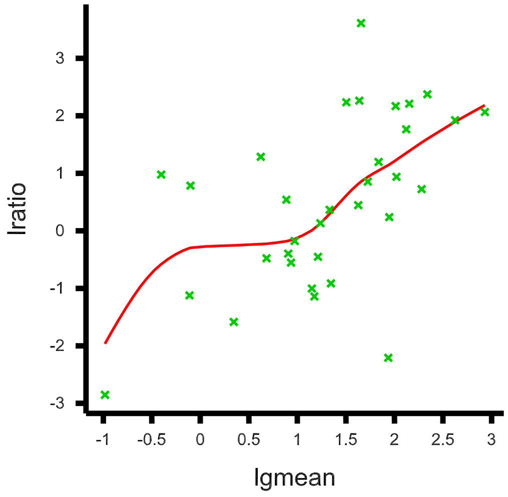

Figure 3.8
HW data: Ratio of case to control-set arithmetic mean exposures against geometric mean of case and control-set mean exposures with smooth fit from GAM

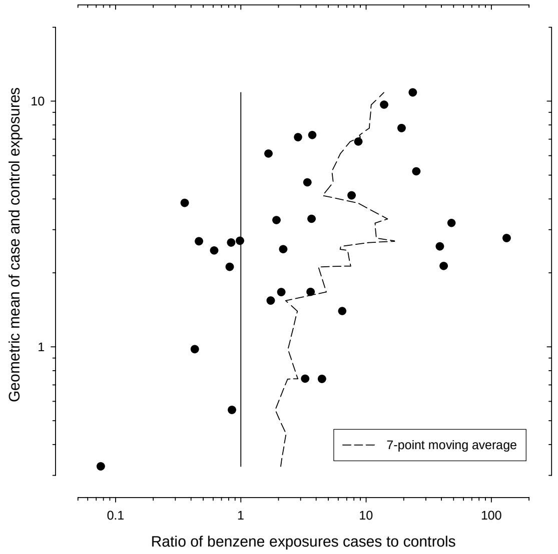

Figure 3.9
HW data: Ratio of case to control-set arithmetic mean exposures against geometric mean of case and control-set geometric mean exposures

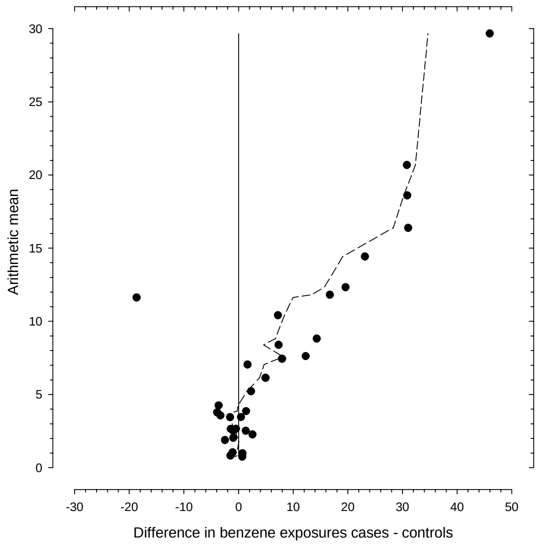

Figure 3.10
HW data: difference between case and control-set arithmetic mean exposures against mean of case and control-set arithmetic mean exposures

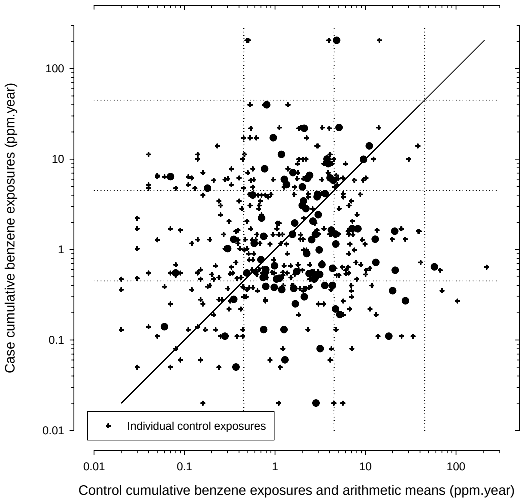

Figure 3.11
IP data: Cumulative benzene exposures of cases against control individual exposures and control-set arithmetic means

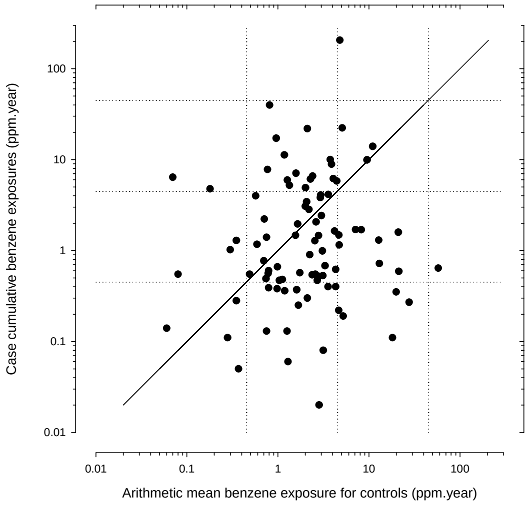

Figure 3.12
IP data: Cumulative benzene exposures of cases against control-set arithmetic means

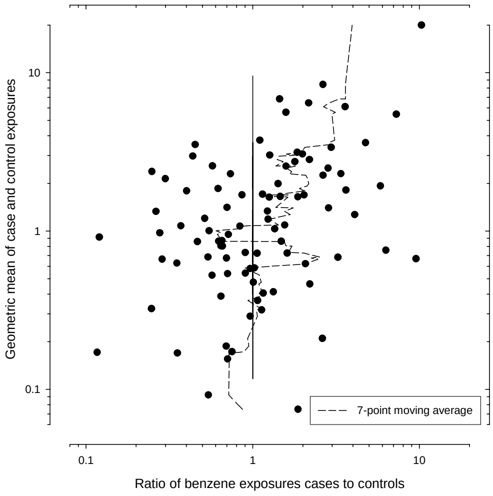

Figure 3.13
IP data: Ratio of case to control-set arithmetic mean exposures against geometric mean of case and control-set arithmetic mean exposures

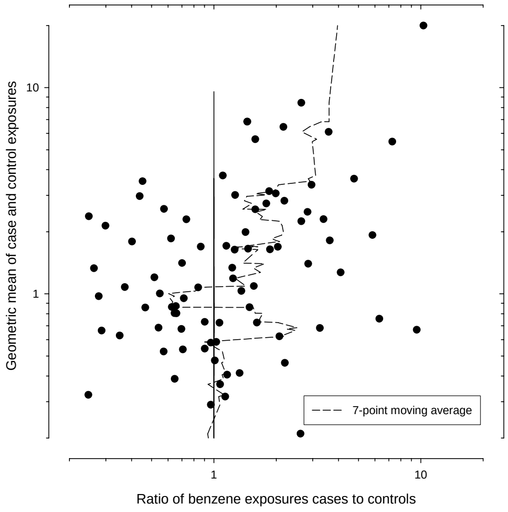

Figure 3.14
IP data: Ratio of case to control-set arithmetic mean exposures against geometric mean of case and control-set geometric mean exposures

3.10 CATALOGUES OF DATA

In our proposal, we envisaged compiling a catalogue of the data held at each of the visited sites, to list the files held and their contents. This would cover original measurements and final calculated values (such as cumulative exposures). Each already had fairly extensive and detailed documentation of their data files and structures, and as a result we judged that cataloguing these afresh would be too time- and labour-consuming, and would represent considerable duplication of previous effort. As a result, we satisfied ourselves that the documentation was indeed in good order, and decided to abandon the objective of compiling fresh catalogues.

4 COMPARABILITY OF DATA AND RESULTS

4.1 COMPARABILITY OF POPULATIONS

Table 4.1 and 4.2 summarise and compare many aspects of the studies.

We note that the populations are all derived from the same industry, and are entirely male. They are all from the same industry, although the structure of the industry and the inclusion criteria differed somewhat across the studies, as regards which sectors were included and whether the cohort included only blue-collar workers. In particular, 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 cases.

For the three studies with benzene exposures, we have shown that there is a difference in the relationship between the person-years of the follow-up and the data used to create the quantitative estimates of benzene exposure concentrations; thus the IOL and IP studies required the greatest degree of extrapolation backwards in time, and the HW study the least. We expect the contribution to the uncertainty in the benzene estimates from extrapolation to periods when there were no measurements to be greatest in the IOL and IP studies, and least in the HW study.

4.2 COMPARABILITY OF CASE DIAGNOSES

In all the studies, the outcomes were classified according to either the 8th or 9th revision of the International Classification of Diseases. There are differences between these, but for the leukaemias the differences are not important, being simply that ICD 8 code 207 “Other and unspecified leukaemia” corresponds to ICD 9 codes 207 “Other specified leukaemia” and 208 “Leukaemia of unspecified cell type”. Both revisions label lymphoid leukaemia as 204, myeloid leukaemia as 205, and monocytic leukaemia as 206. We therefore believe that for discrimination between leukaemia and other diseases, which underlies the definition of cases, it is immaterial whether 8th or 9th ICD revision is employed. (We note that the recent revision to ICD 10 brings a completely different structure to the classification, and epidemiologists are already concerned that creating correspondences in between ICD 10 and the previous versions may be difficult. While this may be important for future studies or extensions of those considered here, it does not affect the diagnoses held in the present studies.)

The definition of casehood differed between studies, in that the IOL and API studies included only deaths, but the IP and HW studies included deaths and cancer registrations. In the latter studies, cases were identified that had been diagnosed and registered with leukaemia, but had not died, either because of successful treatment or because the disease had not yet progressed to death. Of course, diagnosis takes place earlier than death, and the date of definition for a registered case is therefore earlier than if the case had been identified only at death. This in turn affects the age of potential controls. Exposures calculated to an earlier date will of course also be smaller, and thus the distribution of exposures will also have been affected. 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 derived solely from cancer registration alone is small. Clinical experience suggests that, although 30%-50% of AMLs may be cured or remitted by modern treatments, the proportion was much smaller in the past. In addition, it is well known in clinical circles that secondary ANLLs following either chemical exposure or previous chemotherapy have a much poorer prognosis than *de novo* cases. Taken together, these factors suggest that none of the studies is likely to have missed identifying significant numbers of cases, either in the studies that did not include cancer registries or in the early periods when they did not exist, particularly if benzene were involved in their causation.

Analyses that consider separately different cell types have been carried out in the IP and HW studies, and collection of cell type data has begun in the IOL study. It is not clear what cell type information is available for the API study. To date, analyses by cell type have been acknowledged to be inconclusive, largely because they treat subsets of already small numbers of cases. In the IP study, and where diagnosis was uncertain in the HW study, the available medical evidence on cell type was reviewed by a study expert. We cannot tell whether the experts will have applied the same standards of judgment, although it seems unlikely that there would be much disagreement on distinctions between acute lymphoid and acute myeloid conditions. Nowadays, these distinctions are made very reliably by flow cytometry, and previously by cytochemical staining, which was time-consuming but also reliable. Thus, while it is possible that some misclassification might have occurred, the numbers involved would have been very small. However, changes in classifications of just a few cases could have a relatively large impact. Therefore, for any future work involving a breakdown of case by cell types, we believe there would be some advantage in standardising such distinctions as much as possible.

4.3 COMPARABILITY OF CONTROL SELECTION

There were small differences in the way that controls were selected. In all the studies, the controls were same sex. In the IOL study, the cases were stratified by decade of birth, and controls selected from the same strata. In the other two studies, the controls were matched to individual cases, in the IP study without replacement and in the HW with replacement. While the latter is more proper (Clayton and Hills, 1993), the probability of a control being selected more than once will have been small, and therefore the difference will not have had any impact. The difference in control-to-case ratios of 4:1 or 5:1 is not a significant factor in the designs.

4.4 COMPARABILITY OF WORK HISTORY DATA

All the study teams appreciated the need to compile the best possible work histories, within practical constraints, for the case control studies, and all take great pains to achieve this. Necessarily, they worked with what was available. The IOL study got its work histories direct from the company personnel records, and these were relatively complete, with few missing periods. The IP study also used personnel records from their four constituent companies, but there was some missing data, and some supplementation from other sources, including pension records, medical records, and interviews with retired or long service staff. The HW study differed in that material on work histories was given by individuals at interview when they were recruited to the Health Watch cohort study (and later cross-checked with company records). In all of these cases, the histories were based on data that predated the case definitions, and thus would not have been subject to a leukaemia-related reporting bias. Most importantly, we found nothing to suggest that there were likely systematic differences between the reliability or completeness of the work histories of cases and their matched controls.

Table 4.1 Comparative characteristics of parent cohort studies.

	STUDY			
	IOL (Schnatter)	**IP (Rushton)**	**HW (Gun/ Glass)**	**API (Wong)**
Cohort study				
Industry segments	Marketing/ distribution	UK oil distribution centres	Marketing, distribution, upstream, refining	Land-based and marine distribution workers, potentially exposed to gasoline
Number of sites	226	?	235	30+ record locations
Source of ID records	Computerised employee relations database	Personnel and pensions records	Company records	Company records
Inclusion criteria	1) All active empl and annuitants 01/01/1964 2) All new regular empl hired 01/01/1964 – 31/12/1983	Employed 1+ yr 01/01/1950 – 31/12/1975	Employed 5+ yrs since 1980	Employed 1+ yr 1946 – 1985
Exclusions	No females	No females		Units with incomplete records
Cohort size	6,672	23,306	16,252 male, 1273 female	9026 land-based
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)	01/01/1947 – 31/12/1989
#Deaths	1,154	8,743	883 + 520 incident cancers	2066

Table 4.2 Comparative characteristics of case-control studies.

	STUDY			
	IOL (Schnatter)	**IP (Rushton)**	**HW (Gun/ Glass)**	**API (Wong)**
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.	From land-based sites, all cases with leukaemia, MM or kidney cancer as a primary or contributing cause
#Cases	31 (16 Leukaemia, 7 MM, 8 NHL)	91 Leukaemia	79 (33 Leukaemia, 15 MM, 31 NHL)	35 Leukaemia, 11 MM, 12 Kidney Cancer
Case type(s)	Decedent	Decedent and Incident	Decedent and incident	Decedent
Matching criteria	Decade of birth, alive at case ident.	Company, age, alive at case ident.	Year of birth, alive at case death/ regn.	Company, date of birth +- 2 yrs, alive at case death, sex
#Controls per case	4	4	5	5
Leukaemia subtypes analysed	Leukaemia, NHL, MM	ALL, CLL, AML, CML	ALL, CLL, AML, CML, NHL, MM	AML, MM
Source of work histories	Electronic & hard copy personnel records	Personnel records, pension fund, medical records, interviews	Computerised job history, interview of case or proxy	Hard copy personnel records
Benzene exposures considered	CE (ppm-years), mean ppm, maximum intensity, dermal exposure potential	Ppm, CE (ppm-years) peaks, dermal exposure	Ppm, duration, CE (ppm-years)	(Total hydrocarbons) duration, CE, frequency peak exp, time 1st exp
Lags considered	0, 5, 10, 15 yrs	0, 5, 10 yrs	0, 5, 10, 15	
Confounders adjusted for	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	Start date

4.5 COMPARABILITY OF EXPOSURE ASSESSMENT METHODOLOGY

The three benzene case-control studies (IOL, IP, and HW) used basically a similar exposure assessment methodology. All three studies used base estimates of benzene exposure based on quantitative measurement results, which were subsequently adjusted for several modifying factors (so called K-factors) to result in a time-weighted workplace exposure estimate for each work history line. These were multiplied by the time spent in that task/job and summed to result in a subject's cumulative exposure estimate in ppm-years. Because the three studies (IOL, IP, and HW) used a similar exposure assessment methodology, the exposure metrics (ppm.years) are similar too, although the base estimates and extent of extrapolation (indicated by the use of modifying factors) differ considerably (Table 4.3). As a result, we cannot exclude the possibility of systematic differences between the three exposure assessment approaches in the three studies.

The API study used a similar exposure assessment approach to the other case-control studies, but the exposure estimates were for total hydrocarbon (THC) exposure only and not specifically for benzene exposure. This means that if the API study were to be included in a pooled analysis, the exposure assessment would have to be redone.

In addition to the estimates for inhalation exposure, the IOL and IP studies assigned a grade for dermal exposure potential to each work history line (IOL: 1, 2, or 3; IP: none, low, medium, or high) by an expert panel of industrial hygienists. The HW and API studies decided not to use such a dermal exposure scoring system (Table 4.3).

Table 4.3
Comparability of exposure assessment and assignment

	IOL	IP	HW
Exposure period estimated	1909-1983	1909-1993	1941-1998
Exposure algorithm	$WE^b=BE^a*K_1*K_2*\ldots*K_n$	$WE^b=BE^a*K_1*K_2*\ldots*K_n$	Similar, but with extra task-based assignment level
Available input data for Base estimates	340 measurements: - Both benzene and total hydrocarbon - Both short term and full shift	?	>3870 measurements - Mainly task-based measurements
Percentage of person years overlapped with available measurement data	9.9%	9.1%	45.5%
K-factor assignment	K-factor assigned to each work history line (except for background levels)	K-factor assigned to each work history line (except for background levels)[c]	K-factor assigned to 6% of work history lines
Dermal exposure	1, 2, or 3	None, low, medium, high	Not used
Peak exposure	Not used	Intermittent peaks defined based on frequency and duration	Subjects with likelihood of episodic exposures were identified

a BE = base estimate
b WE = workplace exposure
c estimated based on available exposure assessment data files

The IP study further classified exposures qualitatively into 12 categories according to the likelihood whether exposure took place in intermittent peaks, defined by: frequency (daily, weekly, monthly), intensity (1-3 ppm, >3 ppm), and duration (1-15 minutes, 15-60 minutes). The HW study identified subjects with a likelihood of episodic exposures including: handling concentrated benzene or BTX, drum fillers not using local exhaust ventilation, workers in quality control laboratories with poor ventilation, and barge workers handling gasoline or other high benzene content products. The IOL study did not identify peak exposures to benzene (Table 4.3). The API study classified peak exposure, defined as an episode of exposure in excess of 500 ppm for 15-90 minutes, but this was obviously only for total hydrocarbon (THC).

4.6 COMPARABILITY OF INPUT DATA FOR EXPOSURE ASSESSMENT

Although the basic exposure assessment methodology was similar, an evolution of the exposure assessment methodology over time is apparent throughout the three case-control studies, which is mainly evident in the structure and detail of the data that were used to derive base estimates and to assign modifying factors (K-factors).

As illustrated in the Figures on the distribution of person years and overlap in available measurement data in section 3.7, more extrapolation was needed for the IOL and IP studies compared to the HW study, which is reflected in the number of modifying factors (K-factors) used. In the IOL study for all work history lines (except for the population background and site background levels), the exposure estimates were adjusted by one or more K-factors. From the exposure assessment data files that were available for the IP study, it appeared that the majority of work history lines in the IP study were adjusted by one or more K-factors. The exposure assessment report for the HW study shows that only 111 out of 1771 (=6%) task estimates had to be adjusted for one of the expert based modifying K-factors, which confirms that less extrapolation (to adjust for deviations from the base estimate situation) was needed in the HW study compared to the IOL and IP studies (Table 4.3).

The basis for the majority of these modifying factors was expert judgement and K-factors were not validated with exposure measurement data. It was, for instance, not validated that a worker who handled benzol (30% benzene content) had a 10 times higher exposure to benzene compared to a worker who handled gasoline with 3% benzene content.

Although it is hard to quantify the impact of the far reaching extrapolation for exposure estimates to early time periods (for IOL and IP) and the difference in use of K-factors to adjust for deviations from the base estimate situation between studies, we believe that the retrospective exposure estimates of the HW study will be more accurate and reliable than those of the IOL and IP studies.

4.7 COMPARABILITY OF CUMULATIVE EXPOSURE ESTIMATES

Given the comparability in exposure metrics (benzene in ppm-years), it is possible and relevant to compare the cumulative exposure estimates for the three case-control studies (Table 4.4; estimated from available publications, reports and data files). This table illustrates that a relatively large group of subjects in the IOL study were exposed to very low levels of benzene (41% of the population had only background exposure) and that only few subjects were exposed to high levels of benzene (15% of all subjects had a full work history of exposure to benzene) compared to the other two case-control studies. The HW study exists of mainly exposed workers, but seems not to have subjects with high exposure to benzene (maximum is 50.9 ppm.years) The likely explanation for this is that the cohort is younger (started from 1941) and highly exposed jobs did not exist in this study. The IOL and IP studies do include individuals with very high exposures from the earlier periods in the cohort (pre-1940). The comparison in table 4.4 shows that the cumulative exposure estimates are to a large extent influenced by the definitions of the cohorts.

Table 4.4
Cumulative exposure distribution for each case-control study in ppm.years.

	Minimum	25%-perc.	median	75%-perc.	maximum
IOL	0.015	0.217	0.449	7.65	219.8
IP	0.003	0.38	0.95	3.85	218.8
HW	0.005	0.6	2.4	7	50.9

Despite these differences expressed above, we assume 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 systematic differences between studies. Such potential differences cannot be obtained without a more detailed comparison and validation study.

4.8 COMPARABILITY OF POTENTIAL CONFOUNDERS

The only other exposure that was likely to have any impact on the study results was tobacco smoking. All of the studies obtained their data from existing records, 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 14 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 which case its omission as an explanatory variable may have little or no effect. According to Rushton and Romaniuk (1997), reviews of the topic have suggested, at most, a mild increase in risk for smokers. This accords with the current clinical view that smoking is not an important aetiological factor for leukaemia.

4.9 COMPARABILITY OF STATISTICAL METHODS

It is well known that, in case-control studies designed with matching, the statistical analysis needs to reflect the matching. If this is ignored, biased estimates of relative risk may be obtained, and the bias is worst in the case of individual 1:1 or n:1 matching. Appropriate analyses are based on conditional probability arguments and include the Mantel-Haenszel procedure and, more generally, logistic regression techniques fitted by conditional likelihood methods. Some variants and approximations are sometimes used, e.g. fitting terms for group stratification through unconditional likelihood models. However, these are simply different ways of allowing for the matching, and are logically equivalent.

All the studies used conditional regression techniques in their analyses, in some cases including continuous exposure variables, but more often categorising the exposures into groups and calculating relative risks against the lowest-exposed group. The IOL group report that their study involved group-matching on age by decade rather than individual matching, but this variation was not important, and the analysis was, appropriately, carried out using conditional methods. It is our judgment that all the studies analysed their data using appropriate analyses to allow for their case-control designs.

We expect, therefore, that the results of the statistical analyses from each of the studies are comparable. However, one problem arises in that exposure categories in each study have been defined differently. This can be overcome by expressing the results graphically. However, a secondary problem is that the "baseline" category of lowest exposed individuals is defined differently between studies. Since the risks in higher exposure categories are described relative to these different baselines, some of which are relatively poorly determined, it follows that it is 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 pmm-year, had odds ratios of 3.9, 6.1, 2.4 and 5.9. The two highest of these were 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. D Glass, in analyses that have been presented at conferences but not yet published, has shown that different choices of group limits alter the judgment on whether there are significantly increased risks in the middle exposure categories, but that the finding of an increased risk in the highest exposures is robust to these redefinitions. In addition, she has applied standard reference rates to the cohort data and calculated that, in the bottom exposure category, we might expect some seven leukaemia cases rather than the three observed. This lends further weight to the suggestion (see 3.9.3) that the baseline category is an artificially low reference level, and with such small numbers this could easily have arisen by chance. A plausible synthesis of these results is that there is little convincing evidence of any increase in risk below 16 ppm-years, but that it is possible that exposures above this level carry an increased leukaemia risk.

4.10 A COMPARISON OF RESULTS FROM THE STUDIES

While there are many differences in detail between the subjects and the methods used in the studies, we have not identified any important differences in procedures that might have led to different results, and we believe that the results from the studies are broadly comparable from a methodological point of view. This is qualified by the observation that comparison of relative risks from categorised exposures is made difficult by differing baselines (see 4.6). In addition, since we cannot be sure that the k-factors in the exposure assessments were comparable across studies, we have to allow the possibility that the cumulative exposures used, while correctly ranking individuals within each study, may not scale exactly across the studies. This implies that we have greater confidence in statements about qualitative results than in the exact quantification of risks at different exposures, and that we have some uncertainty about the comparability across studies of the cumulative exposures estimated.

We may summarise the overall findings of the three benzene case-control studies as follows:

- The IOL study found no association between leukaemia and benzene exposure
- The IP study found no evidence of a link with benzene for lymphatic leukaemia, and some suggestion of a relationship for myeloid leukaemia
- The HW study showed increased risks of leukaemia with increasing benzene exposures, and identified nonlymphocytic (i.e. essentially myeloid) leukaemia as increased in the highest exposed workers.

On the face of it, these are conflicting, but we have already noted that the IOL study had a large proportion of individuals with background exposure, and therefore very little power to demonstrate a relationship with benzene. The IP study, with its much greater numbers of cases and relatively higher exposures, was able to analyse separately for cell types, and identified myeloid leukaemias as possibly related to benzene, as would be expected from other studies. Finally, the HW study, with its more recent work histories and somewhat larger numbers of exposed workers, showed a clear suggestion of raised leukaemia risks at higher benzene exposures, but had too few cases for meaningful analyses by cell type. These findings are not necessarily inconsistent with each other, nor with results from other studies.

One HW finding that has generated much discussion is the claim that risks are increased at exposures above 2 ppm-years (Glass *et al*, 2003). This conclusion was based on an analysis that compared risks across sub-groups. Our graphical investigations, and other analyses by the HW group (e.g. Glass *et al* in Institute of Petroleum, 2003) show that this conclusion is not robust to different choices of baseline, and that the comparison is exaggerated by a deficit of cases in the lowest exposure group. However, the evidence for an effect at higher exposures, say >10 ppm.years, appears much more robust.

5 OPTIONS FOR POOLING DATA SETS

Since we have been unable to investigate the details of the API study, since it is not known who currently acts as custodian of the data and study materials, and since the study did not quantify benzene exposure *per se*, we do not here consider the API study as a candidate for inclusion in a combined analysis.

5.1 A PREVIOUS REPORT ON COMPARABILITY AND POOLING

The principal scientists from the three benzene case-control studies have already considered many questions of comparability and feasibility of combined analyses (Schnatter *et al*, 2002). They noted the commonalities between the studies, and declared that the exposures were comparable in scale as well as in units. They recommended that future analyses should be based on data for individual subjects, rather than on a meta-analysis of published summaries. They also recommended 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.

We agree with the thrust of many of the recommendations made, but would in some cases go further.

5.2 ADVANTAGES OF A POOLED ANALYSIS

We believe that a simple meta-analysis of the published data summaries is not possible, principally because the exposure categories used to group the quantitative exposures differ. Since the data sets exist at an individual level, a combined analysis should be carried out on a pooled individual data set. This would have the advantage that alternative exposure metrics, e.g. to focus on peak exposures or to account for latency by introducing lags, could be calculated on all subjects in the same way. We note that the ranges of exposures in the studies are complementary rather than congruent, with the IOL study providing a higher proportion of workers in the occupations with lower benzene exposures, and the other studies with relatively more directly exposed workers. Thus each study would add strength differentially to the different range of exposures. This implies in turn that the results of a pooled analysis would not necessarily be a straightforward weighted average of the published summary results, since the relative weightings would not be constant across exposure ranges.

We believe that any analysis should be carried out by a team composed of the PIs from the contributing study, who are most familiar with both the industry and the material of the studies.

5.3 ADDITIONAL WORK TO IMPROVE A POOLED ANALYSIS

We recommend that consideration should be given to number of steps that could enhance a combined analysis. Each of these would have cost implications, and would need to be budgeted carefully.

5.3.1 Extended follow-up for additional cases

While it would be possible to pool the data sets that currently exist, we believe that consideration should be given to increasing the number of cases. The IOL study has the smallest number of cases and the earliest follow-up cutoff date, and so has the greatest scope for extension. The IP cohort study has already had a more recent follow-up, which should yield a reasonable number of new cases: T Sorahan should be able to advise. The HW study was the most recent, and extended follow-up will yield few new cases (at the time of the visits, only three potential new cases had been identified in the cohort).

The recruitment of new cases would have important cost implications, since it would be necessary to collate work histories for new cases and their controls, and exposure assessments might be necessary for new lines of work history. It is likely, however, that existing elements of previous exposure assessments could be reused in many cases.

5.3.2 Classification of cases by cell type

With or without new cases, an important advantage of a pooled analysis would be to add power to analyses of specific cell types. The information to do this is now available for all three studies. There would be an obvious advantage in ensuring that the division of cases be carried out uniformly for all the studies, and to that end we would recommend that the information on cell type for all cases be reviewed centrally by one or more expert haematologists, to agree a final standardised classification. It would also be useful and necessary to standardise the terminology for these classifications to best contemporary usage.

5.3.3 Exposure assessment and assignment

For the three benzene case-control studies (IOL, IP, and HW) exposures are in comparable metrics and are expressed as ppm-years for benzene exposure.

Although pooling of studies would be possible, because the approaches used to produce exposure estimates are comparable and exposure estimates have been expressed in the same metrics, this survey has made clear that systematic differences may exist in terms of exposure assignment (task or job title basis and algorithms to calculate average exposures), assumptions made, amount of extrapolation of exposure estimates into the past, use of expert opinions and modifying factors and treatment and analysis of the data. Therefore, we advise that a pooled analysis should be conducted only after validation and detailed comparison of the crude data available. One issue is the use of expert opinion at different levels of the exposure assessment and assignment. A particular limitation of the approach used so far is that the use of expert evaluations has not been made explicit. This is seen for example in the allocation of the so-called K-factors. Some of these factors seem to have a physical or an empirical basis, but they have not been validated, and their influence on the exposure estimates has not been analysed directly. Thus, the above limitations hamper direct pooling of exposure data from the three studies, and the assignment of K-factors by expert knowledge should be clarified and structured across studies. In addition, we believe that some opportunities to derive K-factors directly from measurement data have not been taken. This approach would make more efficient use of the available exposure data, lead to a more structured estimation procedure and provide information on the variability in the derived factors. The information obtained on variability could then be used as input for a sensitivity analysis to evaluate the robustness of estimates for the exposure-response relation.

We therefore recommend the following for consideration:

1. For a combined analysis, consideration should be given to reworking parts of the exposure assessment and assignment with the objectives of improving integration and standardisation. It would be an advantage if this were done by a small panel of experts with knowledge of the industry and of the studies.
2. As an alternative approach we recommend that data from the available exposure measurements should be pooled, into an international exposure database, and analysed for trends in exposure over time for job titles and tasks present in the cohorts. The statistical models obtained could then be used to produce improved estimates of exposure to benzene and could be complemented with expert judgements where necessary.

3. Information on variability in the estimates should be used as input for sensitivity analyses at the level of exposure-response modelling; and could further be used in sensitivity analyses of the exposure-response relations.

6 CONCLUSIONS

We have reviewed four studies, visiting and inspecting the IOL, IP and HW studies at their sites. We have also reviewed the written descriptions of the API study, but have not been able to view any files or documentation for this study.

At all the sites visited, i.e. for the IOL, IP and HW studies, we received warm and helpful co-operation from the PIs and their staff. We found them knowledgeable about the processes carried out, and it was clear in all cases that their attitudes to their studies were fully professional, diligent and oriented towards maintaining quality in data and results. We judge that all three studies have been carried out carefully and to a high standard of professionalism.

We found the studies to have been generally well conducted. The study designs, selection of cases and controls were appropriate, and care was taken to maximise case ascertainment and to minimise selection bias. We have not discovered any aspect of the study procedures, data collection, processing or analysis that would invalidate any of the individual study results. The comments below on particular aspects of the studies and their comparability are made within this general context.

All of the case-control studies were based on cohort studies, but in some cases the results have been different between the two types of study. This is not surprising, in that the two types make inherently different comparisons, and therefore answer different questions. The cohort studies have employed SMR analyses, which compare the mortality experience of the cohort with that of a general reference population. Each of the cohort studies had limited information on individual exposures to benzene, and therefore did not lend themselves to individual comparisons of risks from benzene exposure. The latter need could be met only by the more precise comparisons of a study that characterised and differentiated benzene exposure between individual cohort members, and this was efficiently done in a case-control study.

Since they were designed to answer different questions, we find it unremarkable that the cohort and case-control studies appear in places to give partially different answers. The results of SMR analyses are notoriously difficult to interpret, principally because of healthy worker effects, whereby an employed population will have a lower mortality than the general population of the same age and sex. Analyses that are based on comparisons within a workforce, as in the present case-control studies, largely avoid this problem. We believe, therefore, that the case-control studies give a much more useful answer to questions of risk from benzene exposure.

No study is or can be perfect, and a number of problems have been noted in each of the studies, such as missing data and the possibility of incomplete case ascertainment. In many cases, we have been able to discover that the extent of such problems was small. In addition, many of these would have had much greater impact on the cohort study results. For example, incomplete case ascertainment may bias downwards SMR estimates, but its effect on a nested case-control study will be a slight reduction in power rather than a bias in the exposure-response relationship, except in cases where strong exposure-specific selection had taken place. We found no evidence of such selective effects.

We were able to inspect much of the material from the studies on site, but did not attempt to follow audit trails for all aspects of the data preparation. In each case, this was because the study materials either originated at points other than their current custodial sites, or because material had been archived or returned to source after processing.

In the case of the IP study, we discovered that the transfer of study materials from Nottingham University to Birmingham University had inadvertently not been completed, and that many files from a PC identified as “Machine B” were now not available, except for a few programs filed in manuscript form. The materials missing included the details of the exposure assessment, and all the data preparation and statistical analysis routines. The previous PI (Lesley Rushton, who left Nottingham University before the handover was completed) has requested assistance from Nottingham University staff to recover these files from archive tapes held there, but as this report was being finalised (April 2005) we had not been informed that that process was complete. We recommend that those efforts should continue. The objective should be to identify and transfer to Birmingham all the files in the Machine B catalogue, which is held by Birmingham, and to document clearly any files in that catalogue that can now not be found. It is not clear at present who is taking the principal responsibility to ensure that the process completes.

Work history data are crucial to any occupational study, and each of the three studies had taken care to obtain as full a work history as possible. Some aspects of work history had to be filled by imputation, but the assumptions made appear to have been sensible, and it was possible in sensitivity analyses to omit the cases affected.

The exposure assessments for the three studies were carried out using essentially the same methodology and approach, and the assessors shared experience and strategic thinking to maximise the comparability of their approaches. The methods employed were refined over time, so that the IP study made fewer assumptions than the IOL study of equality of concentrations in similar jobs across different sites, and the HW study took the exposure assessments to the level of task rather than job. In each study, the assessment was carried out, blind to case/control status, for individual dated lines of work history. This may appear different from the more usual process of constructing a job-exposure matrix (JEM), but the principal difference is in the way the data are organised and stored; by work history line rather than by job or task. Indeed, there is a valid argument for working in this way for a small nested case-control study, where there may be relatively little duplication of jobs or tasks across the cases and controls.

The assessment method was based on assigning “base estimates” of airborne benzene concentrations to typical jobs (or tasks), and applying multiplicative “k-factors” to modify these for different methods of working, input materials, periods, etc. The base estimates were largely derived from measurement data, and appear to be reasonably consistent across the studies; but the k-factors were mostly derived from expert judgment. As a result, we judge that the assessment procedures are likely to have produced exposures that correctly distinguished relative intensities of exposure within each study. However, we have found it difficult to judge whether the k-factors are comparable across the different studies. Greater standardisation and validation of the k-factors might lead to some alterations of the scaling of exposures across the studies, although we cannot know unless such an exercise is carried out.

The statistical analyses carried out have followed standard methodologies, and have been carried out correctly. However, the methods used have some inherent problems, particularly in comparing categories of grouped exposure to a baseline of low exposure. These difficulties are principally in the variability in the baseline level itself, and the problems of multiple comparisons involved in comparing other groups with the baseline as if the comparisons were independent of each other and of the implied increases of dose. These difficulties have complicated comparisons of results across the studies, but comparisons that take account of the different ranges of exposures underlying each study suggest that the studies may be less contradictory than some commentators have suggested, and indeed may be consistent in that the IP and HW studies suggest some increased risk among the highest exposed subjects while the IOL study has too little power to detect any effect of exposure.

It is possible to over-interpret these findings, and with the relatively small numbers of leukaemia cases in each study the effects of chance variation are greater than in studies of more common diseases. We have noted earlier that chance variations are likely to be responsible, in the HW study, for the low number of cases at the very lowest defined exposure group; and that, if this is indeed just chance variation (rather than, say, a protective effect), the apparent increased risks in some exposures between 1 and 16 ppm-years, although formally statistically significant, could simply be chance artefacts. However, we think that chance is an unlikely explanation for the more greatly increased risks in the highest exposure group in that study. In the end of the day, the antidote to chance variation is greater numbers of case-control groups, and that can be achieved here only by pooling the studies.

We believe that the studies are sufficiently similar in scope in methodology to be pooled for a more powerful analysis. We recommend that such an analysis, if it went ahead, should be based on data at the level of individual subjects rather than summary statistics.

We believe that there are a number of ways in which the data could be extended or enhanced prior to a combined analysis, and recommend that these should be considered carefully, balancing practicability with cost. These include:

1. Using the extended follow-up now available to increase the numbers of case-sets in the data base, particularly for the IOL and IP studies;

2. Using all existing measurement data on airborne benzene concentrations to validate the exposure k-factors, and rationalise these across the studies;

3. Setting up a small panel of experts in haematology to rationalise the distinction between leukaemia types across studies;

4. Analysing the data using statistical methods as little influenced by grouping strategies as possible.

6.1 SUMMARY

1. We received willing co-operation from the study teams, and access to whatever study materials we requested that they could find.

2. 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*.

3. The studies were generally well conducted and the conclusions were generally sound (but see 4). In particular, while there were some differences in methods between studies, it is our judgement that they are not responsible – individually or together – for the apparent differences in results between the studies.

4. What seems to be the chance occurrence of very few cases of leukaemia among the lowest-exposed members of the HW study contributed to the perception that there was a positive relationship between exposure to benzene and leukaemia in that study. Independently and in parallel, the original investigators and ourselves concluded that this finding may be spurious. We agree however that the study has shown an effect at higher exposures.

5. The differences between studies did lead to differences in power to detect relationships between benzene and leukaemia. In particular, the IOL study has too

little power to detect any effect of exposure, unless that effect is very large. The negative findings from that study should not weigh strongly in evaluating the evidence overall.

6. On that basis, we think that the apparent discrepancies in results between the studies are most likely to be related to study size and, to an extent, to differences in distributions of exposure. Overall, the three studies are consistent with a view that exposure to benzene increases the risks of developing leukaemia in the higher exposure ranges of the three studies.

7. This view can be tested, and if appropriate the associated relative risks estimated, by pooling data from the three studies.

8. We see no intrinsic barrier to data pooling, other than that some of the core data from the IP study were missing. These data are necessary if core methods are to be re-worked. Work is in progress to find them, and should be continued.

9. There is scope for improvement of methods in some respects. We make concrete suggestions which we propose should be considered if and when a pooled analysis takes place.

10. Doing this study 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.

7 REFERENCES

7.1 PEER REVIEWED LITERATURE AVAILABLE TO THIS REVIEW

7.1.1 IOL study

Armstrong TW, Pearlman ED, Schnatter AR, Bowes SM, 3rd, Murray N, Nicolich MJ. Retrospective benzene and total hydrocarbon exposure assessment for a petroleum marketing and distribution worker epidemiology study. Am Ind Hyg Assoc J. 1996;57:333-343.

Schnatter AR, Acquavella JF, Thompson FS, Donaleski D, Theriault G. (1990). An analysis of death ascertainment and follow-up through Statistics Canada's Mortality Database System. Canadian Journal of Public Health 81(1): 60-65.

Schnatter AR, Armstrong TW, Nicolich MJ, et al. Lymphohaematopoietic malignancies and quantitative estimates of exposure to benzene in Canadian petroleum distribution workers. Occup Environ Med. 1996;53:773-781.

Schnatter AR, Armstrong TW, Thompson LS, et al. The relationship between low-level benzene exposure and leukemia in Canadian petroleum distribution workers. Environ Health Perspect. 1996;104 Suppl 6:1375-1379.

Schnatter AR, Nicolich MJ, Bird MG. Determination of leukemogenic benzene exposure concentrations: refined analyses of the Pliofilm cohort. Risk Anal. 1996;16:833-840.

Schnatter R. Petroleum worker studies and benzene risk assessment. J Toxicol Environ Health A. 2000;61:433-437.

Schnatter AR, Katz AM, Nicolich MJ, Thériault G. A retrospective mortality study among Canadian petroleum marketing and distribution workers. Environ Health Perspect. Supplements 1993;101 (Suppl 6): 85-99.

Schnatter AR, Thériault G, Katz AM, Thompson FS, Donaleski D, Murray N. A retrospective mortality study within operating segments of a petroleum company. Am J Ind Med 1992;22:209-229,

7.1.2 IP study

Alderson M, Rushton L. Mortality patterns in eight U. K. oil refineries. Ann N Y Acad Sci. 1982;381:139-145.

Duarte-Davidson R, Courage C, Rushton L, Levy L. Benzene in the environment: an assessment of the potential risks to the health of the population. Occup Environ Med. 2001;58:2-13.

Lewis SJ, Bell GM, Cordingley N, Pearlman ED, Rushton L. Retrospective estimation of exposure to benzene in a leukaemia case-control study of petroleum marketing and distribution workers in the United Kingdom. Occup Environ Med. 1997;54:167-175.

Rushton L, Alderson MR. A case-control study to investigate the association between exposure to benzene and deaths from leukaemia in oil refinery workers. Br J Cancer. 1981;43:77-84.

Rushton L, Romaniuk H. 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. 1997;54:152-166.

Wong O, Raabe GK. (1998) Acute myeloid and monocytic leukaemia and benzene exposure in petroleum distribution workers in the United Kingdom. Letter. Occup Environ Med; 55: 360-361.

Rushton L, Romaniuk H (1998). Acute myeloid and monocytic leukaemia and benzene exposure in petroleum distribution workers in the United Kingdom. Authors' reply. Occup Environ Med; 55: 361.

Rushton L, Alderson MR. An epidemiological survey of eight oil refineries in Britain. Br J Ind Med. 1981;38:225-234.

Rushton L, Alderson MR. Epidemiological survey of oil distribution centres in Britain. Br J Ind Med. 1983;40:330-339.

Rushton L. A 39-year follow-up of the U.K. oil refinery and distribution center studies: results for kidney cancer and leukemia. Environ Health Perspect. 1993;101 Suppl 6:77-84.

Rushton L. Further follow up of mortality in a United Kingdom oil distribution centre cohort. Br J Ind Med. 1993;50:561-569.

Rushton L. Further follow up of mortality in a United Kingdom oil refinery cohort. Br J Ind Med. 1993;50:549-560.

Rushton L, Romaniuk H. Comparison of the diagnosis of leukaemia from death certificates, cancer registration and histological reports - implications for occupational case-control studies. Br J Cancer. 1997;75:1694-1698.

Rushton L. Benzene exposure in the petroleum distribution industry associated with leukemia in the United Kingdom: overview of the methodology of a case-control study. Environ Health Perspect. 1996;104 Suppl 6:1371-1374.

7.1.3 HW study

Christie D, Robinson K, Gordon I, Rockett I. "Health Watch": the Australian petroleum industry health study surveillance programme. Med J Aust. 1984;141:331-334.

Christie D, Robinson K, Gordon I, Bisby J. A prospective study in the Australian petroleum industry. II. Incidence of cancer. Br J Ind Med. 1991;48:511-514.

Christie D, Robinson K, Gordon I, Bisby J. A prospective study in the Australian petroleum industry. I. Mortality. Br J Ind Med. 1991;48:507-510.

Glass DC, Gray CN, Jolley DJ, et al. Leukemia risk associated with low-level benzene exposure. Epidemiology. 2003;14:569-577.

Glass DC, Adams GG, Manuell RW, Bisby JA. Retrospective exposure assessment for benzene in the Australian petroleum industry. Ann Occup Hyg. 2000;44:301-320.

Glass DC, Gray CN. Estimating mean exposures from censored data: exposure to benzene in the Australian petroleum industry. Ann Occup Hyg. 2001;45:275-282.

Glass DC, Gray CN, Adams GG, Manuell RW, Bisby JA. Validation of exposure estimation for benzene in the Australian petroleum industry. Toxicol Ind Health. 2001;17:113-127.

Glass DC, Sim MR, Fritschi L, Gray CN, Jolley DJ, Gibbons C. Leukemia risk and relevant benzene exposure period-Re: follow-up time on risk estimates, Am J Ind Med 42:481-489, 2002. Am J Ind Med. 2004;45:222-223; author reply 224-225.

Gun RT, Pratt NL, Griffith EC, Adams GG, Bisby JA, Robinson KL. Update of a prospective study of mortality and cancer incidence in the Australian petroleum industry. Occup Environ Med. 2004;61:150-156.

7.1.4 API study

Raabe GK, Wong O. Leukemia mortality by cell type in petroleum workers with potential exposure to benzene. Environ Health Perspect. 1996;104 Suppl 6:1381-1392.

Tsai SP, Wen CP, Weiss NS, Wong O, McClellan WA, Gibson RL. Retrospective mortality and medical surveillance studies of workers in benzene areas of refineries. J Occup Med. 1983;25:685-692.

Wong O, Trent L, Harris F. Nested case-control study of leukaemia, multiple myeloma, and kidney cancer in a cohort of petroleum workers exposed to gasoline. Occup Environ Med. 1999;56:217-221.

Wong O, Raabe GK. Non-Hodgkin's lymphoma and exposure to benzene in a multinational cohort of more than 308,000 petroleum workers, 1937 to 1996. J Occup Environ Med. 2000;42:554-568.

Wong O, Raabe GK. Multiple myeloma and benzene exposure in a multinational cohort of more than 250,000 petroleum workers. Regul Toxicol Pharmacol. 1997;26:188-199.

Wong O. An industry wide mortality study of chemical workers occupationally exposed to benzene. II. Dose response analyses. Br J Ind Med. 1987;44:382-395.

Wong O. An industry wide mortality study of chemical workers occupationally exposed to benzene. I. General results. Br J Ind Med. 1987;44:365-381.

Wong O. Risk of acute myeloid leukaemia and multiple myeloma in workers exposed to benzene. Occup Environ Med. 1995;52:380-384.

Wong O, Raabe GK. Cell-type-specific leukemia analyses in a combined cohort of more than 208,000 petroleum workers in the United States and the United Kingdom, 1937-1989. Regul Toxicol Pharmacol. 1995;21:307-321.

Wong O. A critique of the exposure assessment in the epidemiologic study of benzene-exposed workers in China conducted by the Chinese Academy of Preventive Medicine and the US National Cancer Institute. Regul Toxicol Pharmacol. 1999;30:259-267.

Bergsagel DE, Wong O, Bergsagel PL, et al. Benzene and multiple myeloma: appraisal of the scientific evidence. Blood. 1999;94:1174-1182.

Wong O. Recent findings and new initiatives for epidemiologic research on benzene. J Toxicol Environ Health A. 2000;61:457-466.

Wong O, Raabe GK. Acute myeloid and monocytic leukaemia and benzene exposure in petroleum distribution workers in the United Kingdom. Occup Environ Med. 1998;55:360-361.

Wong O, Raabe GK. Critical review of cancer epidemiology in petroleum industry employees, with a quantitative meta-analysis by cancer site. Am J Ind Med. 1989;15:283-310.

Wong O, Raabe GK. A critical review of cancer epidemiology in the petroleum industry, with a meta-analysis of a combined database of more than 350,000 workers. Regul Toxicol Pharmacol. 2000;32:78-98.

Wong O, Harris F, Smith TJ. Health effects of gasoline exposure. I. Exposure assessment for US distribution workers. Environ Health Perspect. Supplements 1993;101 (Suppl 6): 13-21.

Wong O, Harris F, Smith TJ. Health effects of gasoline exposure. II. Mortality patterns of distribution workers in the United States. Environ Health Perspect. Supplements 1993;101 (Suppl 6): 63-76.

7.2 KEY TECHNICAL REPORTS AVAILABLE TO THIS REVIEW

7.2.1 IOL study

Armstrong, TW. (1991) Exposure assessment protocol for a case control study of benzene related health effects among distribution workers in Imperial Oil Limited. (Draft) Exxon Biomedical Sciences, Inc.

Schnatter AR, Thériault G, Katz AM, Thompson FS, Danaleski D, Murray N. (1991) A retrospective mortality study within operating segments of a petroleum company. Exxon Biomedical Sciences, Inc.

7.2.2 IP study

Rushton L, Alderson MR. (1980). An epidemiological survey of eight oili refineries in the UK. Final report. Institute of Cancer Research.

Bell GM, Cordingley N, Lewis SJ, Pearlman ED. (undated). Retrospective estimation of exposure to benzene in a leukaemia case-control study of petroleum marketing and distribution workers in the united kingdom. Exposure Assessment Task Force.

Rushton L, Romaniuk H. (1995) The Institute of Petroleum epidemiological study. Results of 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. Univeristy of Nottingham.

Hurley F, Cherrie J, Buchanan D. (1995). QA/QC audit of aspects of the IP case-control mortality study on benzene and leukaemia at the University of Nottingham. Institute of Occupational Medicine confidential report to client.

7.2.3 HW study

Health Watch. (1981). The Australian Institute of Petrol Health Watch surveillance program. Annual Report.

Then regular reports until

The University of Adelaide. (2000). Health Watch. The Australian Institute of Petrol Health Watch surveillance program. Eleventh Report.

Glass DC, Adams GG, Manuell RW, Bisby JA. (1998). Retrospective exposure assessment for benzene in the Australian petroleum industry. Executive summary.

Monash University and Deakin University. (2001). Lympho-haematopoietic cancer and exposure to benzene in the Australian petroleum industry. Technical report and appendices.

7.2.4 API study

Wong, O, Morgan, RW. (1985) Feasibility analysis for an historical prospective mortality study of employees exposed to downstream gasoline in the petroleum industry. Environmental Health Associates Inc.

Smith, TJ. (1993) An exposure assessment for marketing and marine distribution workers in the petroleum industry with potential exposure to gasoline. API publication 45552.

Wong, O, Harris, F (1992). A mortality study of marketing and marine distribution workers with potential exposure to gasoline in the petroleum industry. Final report. ENSR Health Sciences.

Wong O, Trent L. (1993). A nested case-control study of kidney cancer, leukemia and multiple myeloma in a cohort of land-based terminal workers exposed to gasoline in the petroleum industry. ENSR Document.

7.2.5 General

Institute of Petroleum (2003). Leukaemia risks in relation to benzene exposure. Report iof a meeting held on 22nd October, 2002.

Schnatter AR, Gray C, Rushton L, Glasss D, Armstrong T. (2002). Comparability and feasibility of enhanced analyses of petroleum industry benzene case-control studies. Exxon Biomedical Sciences, Inc.

7.3 ADDITIONAL REFERENCES

Breslow NE, Day NE. (1980). Statistical methods in cancer research. Volume 1 – The analysis of case-control studies. Lyon: IARC. (Scientific publications NO 32)

Coker DT. (1984). Bodily uptake rate differences between real life fluctuating concentrations and steady state laboratory conditions. XXI International Congress on Occupational Health, Dublin, Ireland.

Clayton D, Hills M. (1993). Statistical models in epidemiology. Oxford: Oxford University Press.

Hastie TJ, Tibshirani RJ. (1990). Generalized Additive Models. London: Chapman & Hall.

Parkin DM, Whelan SL, Ferlay J, Teppo L, Thomas DB. (2002). Cancer incidence in five continents, Volume VIII. Lyon, France: International Agency for Research on Cancer.

Rinsky RA, Smith AB, Hornung R, Filloon TG, Young RJ, Okun AH, Landrigan PJ. (1987). Benzene and leukaemia. An epidemiologic risk assessment. New England J Medicine; 316: 1044-1050.

Seixas NS, Sheppard L and Neitzel R. (2003). Comparison of task-based estimates with full-shift measurements of noise exposure. AIHA J; 64(6): 823-829.

APPENDIX A1. REPORT OF SITE VISIT TO CLINTON, NJ: IOL STUDY

A1.1 INTRODUCTION

This case-control study of Canadian petroleum distribution workers grew out of a cohort study of workers within Imperial Oil Limited (IOL). Because IOL did not have a dedicated in-house epidemiology unit, the study was carried out in conjunction with ExxonMobil Biomedical Sciences, Inc, (EMBSI), now sited at Clinton, New Jersey.

The source of vital status information in this study was the government agency Statistics Canada, and liaison with them was done entirely by IOL, under the coordination of Lorri S Thompson. Under the study protocols, the work history data and related data on individuals that were passed to EMBSI were anonymised, and identified only by a study identifier, and EMBSI do not hold any definitive identifiers of the study subjects. For the protection of confidentiality, the key that associates study number with identity is held only by Statistics Canada.

We had agreed to visit only the EMBSI site at Clinton, NJ, under the stewardship of PI Rob Schnatter. Lorri Thompson also attended, and her descriptions of procedures were invaluable. We were able, during the visit, to inspect a great many archived documents, including the data collection forms that were passed to EMBSI. However, because the separation and anonymising of the data, it was not possible to observe the records from which these data were extracted, nor to audit the process of their extraction. Indeed, we cannot even request source data on any of the subjects without asking Statistics Canada to matching back to individual identities, a procedure that was rejected as too cumbersome.

A1.2 COHORT DEFINITION

The cohort study has developed in stages. In the study known as IOL1, follow-up of around 25k subjects was up to 1973. In IOL2, which followed 34k subjects up to 1983, there were many different operating segments of the industry: refining, upstream, pipeline, marketing & distribution, office, building products, chemicals, coal and minerals, and marine. IOL2 was the base cohort for the case-control study, and the subcohort of marketing and distribution workers. The IOL2 study population consisted of all active employees and living retires on 1 January 1964, and all new regular employees hired between that date and 31 December 1983, of one large Canadian petroleum company, Imperial Oil Limited (IOL). IOL3 brought the cohort up to 42K, with a follow-up to 1994 and a new system, ETHIS, for characterising exposures.

This definition of the study population "was primarily driven by the availability of a computerized employee relations database, which is used for payroll and benefits purposes within the company".

Because the construction of the cohort data files had been performed directly from the IOL payroll and benefits computer database, there were no original hard copy records available for inspection at the Clinton site, and it was therefore not possible to completely audit that process or its completeness. The published study reports note in fact that no formal testing of the completeness of the cohort was carried out, while noting that a database controlling payroll and benefits is expected to exhibit very great completeness. Those involved report having had no reason to doubt that coverage of the desired occupations was complete. They have reported that both computerized and manual quality control and consistency checks were carried out on the data and on cohort eligibility, including a random check of 200 records against hard copy sources, which showed almost complete agreement on dates of birth and employment.

Cause-specific mortality data for cohort members were obtained from Statistics Canada (SC). Cohort members active in 1983 or known to be receiving benefits were noted as alive, and the 60remainder sent for tracing. Other sources for tracing deaths were the company's own database and the US National Death Index. Of the deaths identified, 96% came from SC.

In the analyses of cohort mortality, employees not found to be dead were assumed alive until the end of the study. A published evaluation of this assumption found an under-ascertainment of 2.4% for known deaths in the cohort, suggesting that the SMRs might be a slight underestimate. All the deaths supplied by SC had been coded to ICD codes for underlying cause according to the ICD revision in effect at the time of death. For the small number of deaths not identified through SC, a certified nosologist determined and coded the underlying cause from the death certificate.

Work history data for the cohort were extracted from the computer databases. These comprised location, department, job title, and a code for job function, which was later used in assigning hydrocarbon exposure categories. Work history data had been computerised between 1960 and 1964 at the different locations, but some incompleteness was noted in records up to 1968. For 2529 employees who quit before 1960, only their last line of work history (i.e. last known job) was captured, and was assumed to typify their whole employment (in the cohort study); for those whose date of first employment predated the first line of work history, it was assumed that this represented the earliest work history also. There were on average 4.0 work history segments per employee.

The cohort study was analysed with standard SMR methodology and Poisson regression techniques. For all-cause mortality, the usual healthy-worker effect was found. Tank truck drivers had five leukaemia deaths, giving an SMR of 3.35, but leukaemia mortality was not elevated in the larger group of workers classified as exposed to hydrocarbons. Mortality from multiple myeloma was 1.8 times higher than comparable Canadian national rates.

A1.3 CASE IDENTIFICATION

Given the observations from the cohort study, it was decided to perform a nested case-control study, specifically to provide new data on the risk of leukaemia in benzene workers with chronic low level exposure to benzene. From the data files holding the results of the tracing exercise, cases were identified who had

(a) died with an underlying cause of either leukaemia (ICD8 codes 204-207), multiple myeloma (ICD8 203) or non-Hodgkin lymphoma (ICD8 200, 202.0,202,1, 202.2 or 202.9); (The ICD revision in effect at the time of death was used to define cases)
(b) ever worked in either the marketing or distributions, marine or pipeline segments;
(c) died between 1964 and 1983.

These criteria yielded 16 leukaemias, seven multiple myelomas and eight non-Hodgkin lymphomas; 31 deaths in all. It was not possible at that time to obtain reliable information on the leukaemia cell types. (Since publication of the case control study, cell type information for the cases has been obtained and would be available to future work.)

The study team ran a quality control check of the SC ascertainment, by tracking their performance on 757 deaths previously known. Of those submitted, 93.1% were detected in SC records, including 97.6% of those deaths that occurred in Canada. There were few links to employees presumed alive, indicating 99.8% specificity for these records, at least for deaths detected in Canada through SC.

A1.4 CONTROL IDENTIFICATION

Controls were selected from the cohort in a ratio of four to each case. They were restricted to males (like the cases), frequency matched by decade of birth and alive on or after the cases' dates of death. This yielded 124 matched controls.

A1.5 DATA COLLECTION IN CASES AND CONTROLS

The cohort study had collected data on work histories from the computerised database systems. For all workers with incomplete computerized work histories, the case-control study collected work histories afresh from hard copy personnel records (see below).

Potential confounding variables, where available, were extracted from company medical records, and included information on smoking habits, hobbies, previous occupations and exposures, diagnostic radiation exposure, and family history of cancer. The data were reasonably detailed, although not available for all subjects. Completed abstraction forms were available for these records which allowed comparisons with the computerized data.

No data were collected from the cases and controls themselves, nor from any proxy respondents.

A1.6 WORK HISTORIES IN CASES AND CONTROLS

Staff of IOL extracted data on work histories for the selected cases and controls, from the hard copy personnel records of the company, blind to case/control status. The data were recorded by hand, according to a written protocol, on pre-prepared forms. These had space for recording the worker's identification number and name, and all the jobs held, plus absences or gaps of 30 or more days. For the jobs, the forms recorded start date, location, department and job description.

Jobs were not pre-coded. A description of the job title was used (e.g. barrel filler, driver, gauge mechanic, etc.). Subsequently, job titles were clustered into the job title categories described in table 1 of the exposure paper (Armstrong et al., 1996), as described below.

Where possible, the details of work histories were checked with information in the medical records, because the location of an employee dictated where his medical examinations took place. However, the medical records were much less detailed on work histories, so verification was at best partial.

Each recorded change of job defined what became known as a "line" of work history, because it was entered as a separate record or "line" of the study database.

Twenty-one out of 1086 (=1.9%) work history lines (unique combination of location, department, job title and time period) had missing data in work location, department and/or job title. Initially, no exposure estimates were assigned to these work history lines. Later in the process in the final dataset, these gaps in exposure estimates were filled based on previous and/or subsequent jobs for the same individual ("bridging") by the industrial hygienist. An independent science advisory board also reviewed specific assumptions on bridging. Two of those twenty-one were excluded from the dataset, due to insufficient work history data.

A1.7 EXPOSURE ASSESSMENT AND ASSIGNMENT

The methodology for assigning exposures on the basis of work history data and measured concentrations within the industry sector was designed specifically for this study, by occupational hygienists with knowledge of the industry. While other studies have used the principle of the Job-Exposure Matrix (JEM) to do this, the present study used an alternative approach, where typical jobs and periods were assigned Base Estimates of concentration, and each work history line was assigned a series of modifiers ("K-factors"), assigned by expert knowledge, by which local conditions and practices could modify the base estimates.

This approach has many features in common with the JEM approach, and may be expected to produce more and finer distinctions between elements of an individuals' work history. It can be somewhat less transparent, unless all the judgments made in the process of assigning base estimates and K-factors are documented.

A1.7.1 Background

All of the study participants were from the distribution, marketing and marine sectors of the petroleum industry. Based on blue prints, reports, etc., the industrial hygienists concluded that there was not much variety between similar plants in different locations, but that differences in environmental concentrations were likely at least as much from other sources, such as the actual working practices involved. In addition, regional differences due to supply patterns were important. Therefore, exposure estimates and modifying factors were assigned largely by region (to account for regional refinery supplies) and job category but not necessarily a specific plant within a region.

Some information about important historical changes in plants and in working procedures was reported in the IOL survey reports (between 1953 and present). In addition, a lot of information was extracted from technical reports, blue prints, IOL magazine reports, etc. on historical changes in plants. Questionnaires were sent to retirees (through a society of retirees of IOL) and retirees were interviewed to get an overview of process technology and work practices in the different time periods. No systematic overview about which of these sources of information were used in the process of assigning exposures or modifying factors was made and specifics could not be recalled during the site-visit. However, extensive working notes were kept that, with additional effort to review and consolidate, could form a summary overview of the process applied. Notes in a copy of the estimating spreadsheet (provided to the IOM team) specify comments on the base estimate and modifying factor selection process

A1.7.2 Compiling historical data about the plants

For compiling historical data about the plants, a range of sources were employed:

- IOL industrial hygiene reports
- Questionnaires were sent to retirees
- Personal interviews with retirees
- Interview with John Johnson (first occupational hygienist for IOL since 1953)
- Technical reports
- Blue prints
- Photographs
- IOL magazine articles

A1.7.3 Exposure measurements

Company archives provided approximately 107 industrial hygiene survey reports for Canadian marketing and distribution operations (from 1953 onwards), copies of which were all stored at the Clinton site and could be easily accessed during the site visit. Quantitative full-shift, time-weighted average exposure measurements began in the early 1970s. The most rigorously described and most extensive data (greatest number) were selected for the job categories in the case-control study. Those are listed in table 1 of the exposure assessment paper (Armstrong et al., 1996; see below), which lists a mixture of IOL internal company industrial hygiene survey reports and non-IOL publications like published papers, a CONCAWE report and a CPPI report:

1. Verma DK, Julian JA, Bebee G, Cheng WK, Holborn K, Shaw L (1992). Hydrocarbon exposures at petroleum bulk terminals and agencies. *Am Ind Hyg Assoc J* 53(10):645-656.
2. Saunders GA. Internal company industrial hygiene survey report. August 29-September 2, 1977.
3. Tims JM, Conrad M, Cutolo GF, Dodsworth VA, et al. Exposure to atmospheric benzene vapor associated with motor gasoline. CONCAWE report no. 2/81, Brussels, 1981.
4. Saunders GA. Internal company industrial hygiene survey report. February 14-15, 1978.
5. Stacey RA. Internal company industrial hygiene survey report. August 28-September 1, 1977.
6. Standard Oil Development Co. Internal company industrial hygiene survey report. July 25-31, 1953.
7. Coker DT, Christian F, Clayton MF, Irvine D, et al. A survey of exposures to gasoline vapor. CONCAWE report no. 4/87, Brussels, 1987.
8. Parkinson GS (1971). Benzene in motor gasoline - an investigation into possible health hazards in and around filling stations and in normal transport operations. *Ann Occup Hyg* 14(2):145-153.
9. Cheng WK, Bebee G, Froats JFK, Reczek EB. A study of exposure to motor gasoline hydrocarbons vapors at service stations, summary report. CPPI (PACE) report no. 90-8, Ottawa, Ontario, Canada: CPPI (PACE), October 1990.
10. Conrad EH. Gasoline vapor sampling at marketing terminal perimeter. Paper presented at the American Petroleum Institute Motor Gasoline Symposium, May 27, 1982.
11. Wallace LA (1989). Major sources of benzene exposure. *Environ Health Perspect* 82:165-169.
12. Multiple internal company industrial hygiene survey reports (IOL).

Measurements were collected within the cohort (IOL reports: sources 1, 2, 4, 5, 6, 9, and 12), but also outside the cohort (published papers, CONCAWE report and CPPI report: sources 3, 7, 8, and 11).

Measurements described in the IOL reports (sources 2, 4, 5, 6, 12) were routinely taken and not specifically focussed on marketing and distribution workers. In the Verma paper (source 1), measurements were taken to evaluate occupational exposures to gasoline at bulk terminals and agencies. In the CONCAWE reports by Tims et al. and by Coker et al. (sources 3 and 7), the paper by Parkinson (source 8) and the CPPI report by Cheng et al. (source 9), measurements were taken to establish typical levels of exposure to total hydrocarbon and benzene for personnel involved in gasoline manufacturing and distribution systems. The report by Conrad (source 10) and the paper by Wallace (source 11) present results of measurements taken to estimate background benzene exposure levels (facility background and general population background).

Different sampling strategies were employed at different locations and at different times:

Source 1: Measurements were taken in six Ontario petroleum companies between June and September 1986, using a uniform sampling plan. Jobs at the bulk terminal and agencies that were expected to involve hydrocarbon exposures were to be sampled.

Source 2: Measurements were collected in the IOL Ottawa terminal during the week of August 29-September 2, 1977. Exposure levels were determined by riding with individual drivers over a part of their shift or throughout their entire shift and measuring each significant portion of their daily exposure.

Source 3: Measurements were collected at five CONCAWE companies concentrated on areas and operations within European gasoline manufacturing and distribution systems (manufacturing sources, bulk loading facilities (marine, rail, vehicle), service stations, ambient air in other areas) during 1977 and 1978.

Source 4: Measurements were collected in the barrel filling building of one IOL marketing plant in February 1978.

Source 5: Measurements were collected at an IOL terminal in August/September 1977 to determine the concentrations of gasoline and benzene vapours present during truck loading.

Source 6: Measurements were collected in an IOL plant in July 1953 and were concerned with warehouse operations, tank truck filling and mechanical services.

Source 7: This report presents data on 540 personal gasoline (and individual components) vapour exposure measurements collected in 13 European countries during 1984 and 1985 for job groups involved in the manufacture and distribution of motor gasoline.

Source 8: Measurements were collected at nine filling stations in both the South and North of the United Kingdom in August and September 1969 from filling station operators.

Source 9: Throughout 1985 and 1986, atmospheric sampling was carried out at multiple locations in five Canadian urban centres.

Source 10: Gasoline vapour levels were measured at thirteen sites around the perimeter of a large marketing terminal in May 1982.

Source 11: This paper describes results of population-weighted personal exposures to benzene on six locations in the USA.

Source 12: The multiple internal company industrial hygiene survey reports could not be tracked down, so it is unclear what sampling strategy was used for those surveys.

There were a number of detailed descriptions in the source documents of what measurements (personal, static) had been taken:

Source 1: Each of the six member companies collected 35 (15 long-term and 20 short-term) personal samples using a 600-mg activated charcoal tube. The air-sampling flow rate for long-term measurements was 50 cm^3/min and 1 L/min for short-term measurements. Air samples were analysed for 55 hydrocarbon components (of which one was benzene).

Source 2: Personal air samples of gasoline and benzene vapour were collected during loading and unloading procedures at a variety of locations, using an activated charcoal air sampling tube (NIOSH 150 mg type).

Source 3: A total of seven service station mechanics at two service stations were monitored for benzene exposure during one working day. Measurements were collected over the whole 8-hour working day and 15-min. measurements were taken during tasks during which peak exposures might be expected to occur.

Source 4: Seven personal air samples were collected during barrel filling of various products using activated charcoal tubes (NIOSH 150 mg type) at a flow rate of 100 ml/min. Sampling times ranged between 8 and 84 minutes.

Source 5: Five personal full-shift (510-630 minutes) air samples were collected during loading using activated charcoal tubes (NIOSH type).

Source 6: Fourteen short-term personal air samples were taken during loading and were reported as hydrocarbon vapour (ppm). Benzene exposure levels were derived from hydrocarbon results. Not much is reported on the sampling procedure and the duration of sampling.

Source 7: Thirty-two full-shift personal air measurements were taken and analysed for benzene. Because this CONCAWE report describes a summary of results it is not entirely clear which data have been used to construct the base estimate.

Source 8: 30-min. samples were taken using Casella Mark III personal samplers at a rate of 1 L/min.

Source 9: Both short term and long term (sampling time not reported) exposure to benzene was measured for service station attendants during Summer and Winter periods. From the report it is not clear which results have been used for the base estimate.

Source 10: Stationary measurements were collected around the perimeter of the terminal using passive 3M organic vapour monitors during a six-day continuous period (sampling time: 142-144 continuous hours).

Source 11: Each of the 700 subjects of the target population collected two 12-hr air samples using Tenax cartridges. Concurrent outdoor air samples were collected from the backyards of a subset (about 200)

Source 12: The multiple internal company industrial hygiene survey reports could not be tracked down, so it is unclear what samples were collected.

Background information collected regarding the measurements was as follows:

Source 1: A sampling data sheet was designed, which included: air temperature, barometric pressure, relative humidity, wind speed and direction, product temperature, amount and type of gas and products loaded during sampling, the number of trips taken, whether the worker smoked during the sapling period, and a company code.

Source 2: Information was reported on sampling location, amount of gasoline loaded/unloaded, air temperature and wind speed.

Source 3: No additional background information was reported.

Source 4: The product and the number of barrels filled were reported.

Source 5: Information was collected on the number of loaded trucks, the loading technique, temperature, and wind speed.

Source 6: A description of the workplace was given and the work location and handled product.

Source 7: Information on job details, sampling times, and ambient weather conditions, were reported according to a standardised protocol.

Source 8: Background information was reported on the type and location of the stations, weather conditions (temperature, wind speed), benzene content and gasoline throughput.

Source 9: Information was recorded on the job title (kiosk operator or pumper), season (Summer or Winter), smoking status.

Source 10: A description of the plant was given together with information on the throughput of gasoline per day and weather parameters (temperature, relative humidity, barometric pressure, and wind direction and speed) four times a day during the six-day sampling period.

Source 11: Information was collected on age, sex, occupation, and smoking status for each person in the household.

Source 12: The multiple internal company industrial hygiene survey reports could not be tracked down, so it is unclear what background information was collected for those measurements.

Arithmetic mean exposure levels were calculated from the reports and papers per job category and used as Base Estimates (BE) for that job title (see Table A1.1). It is unclear how base estimates were constructed when multiple sources were used to supply measurement data. The paper by Armstrong et al. (1996) states that "*…the authors preferred not to combine data* *from different survey reports, because these represented a different set of operating variables and environmental conditions".* However, table 1 of that same paper presents several sources used per job category.

Subsequently, those Base Estimates were multiplied by modifying factors (so called 'K-factors') to adjust for deviations from the base estimate situation in:

- workplace (K_W), adjusted for technology in use;
- environment (K_E), adjusted for outside temperature using Raoult's law: $\frac{10^{((-1784.8/(273.8+Temp))+7.962)}}{68.2}$ and for indoor ventilation;
- materials (K_M), adjusted for benzene content, fuel volatility and product mix;
- task (K_T), adjusted for work practices and changes in frequency or duration for specific tasks.

The modifying factors to adjust for workplace (K_W) and task (K_T) were based on several sources of information (listed in section A1.7.2), which were reviewed and used by the industrial hygienists to assign a modifying factor to a certain base estimate based on expert judgement. This process is not fully documented and could not be recalled in detail during the site visit. Some notes on the K factor assignment are provided in the "comments" spreadsheet given to the IOM research team, but these notes do not cover all of the process in full detail. Documentation may still be provided retrospectively by the EMBSI investigators, if requested to do so as part of follow-up work. The modifying factors that were used to adjust for outside temperature (K_E) and benzene content in the product handled (K_M) were derived from more objective measures, because the benzene content in gasoline and the annual average climate data were known.

The multiplication of base estimates by the modifying factors resulted in a workplace exposure (WE) per work history line, which could then be summed for each individual to result in a cumulative exposure estimate (CE) taking into account the years spent in a certain job.

A certainty score (1, 2, or 3) was assigned to each base estimate by the industrial hygienist (1=best effort extrapolation, not much objective data; 2= substantial info from interviews, blue prints, etc. to make an estimate; 3= most certain, good set of direct measurements, not much extrapolation). Certainty score 1 was assigned to 24% of work history lines, score 2 to 36% of work history lines and score 3 to 40% of work history lines. None of these uncertainty factors were used in the subsequent epidemiologic analyses.

A1.7.4 Assigning exposures

Throughout the exposure assignment process, industrial hygienists were blinded to case/control status.

Base estimates were calculated for inhalation exposure only. In addition, a score (1, 2 or 3) was given for "dermal exposure potential" based on job title only (and not on company location, department or time period).

No biomarker data were used.

Base estimates were assigned to each work history line (combination of company, location, department, and job title) based on job title (and sometimes time period, when data were available) as reported in table 1 in the exposure assessment paper (Armstrong et al., 1996). A copy of this table is presented in Table A1.1:

Table A1.1
Industrial Hygiene Source Data for Base Estimates of Exposure and Frequency of Use (Armstrong et al., 1996)

Job title	Base Estimate concentration Benzene (ppm)	Number of Measurements reported for the job title[A]
Plantman, plant mechanic	0.12	23
Fuel delivery driver (agency)	0.22	21
Fuel delivery drivers (modern, large terminal)	0.09[B], 0.14[C]	4, 15
Garage mechanic, pump mechanic	0.15	7
Checker, gauger	0.03	4[D,E]
Barrel filler	0.04	7
Loader (modern)	0.34	5[E], 22
Loader (historic)	2.6[F]	14[E]
Marine deck attendant	0.25	26[D,E]
Marine pumpman	1.1[F]	3[E]
Agent, acting agent	0.55	9[D], 32[D], 23
Air & service station attendants	1.3	121
Barrel washer	8.7[F]	4[E]
Facility background job	0.01	13
General population background	0.005	1335

A	Full shift measurements unless otherwise noted
B	Bottom load
C	Top load
D	Full shift
E	Short term
F	Derived from total hydrocarbon result

As shown in table A1.1, seven out of 14 base estimates were derived from less than 10 measurement data points, and some were based on only 3 or 4 samples (fuel delivery driver, checker/gauger, marine pumpman, and barrel washer). Some BEs were derived from total hydrocarbon exposure results (loader, marine pumpman, and barrel washer) and some were based on both short-term and full shift measurements. It is unclear how these different measurements were combined to construct a BE.

Subsequently adjustment factors (K-factors) were assigned to each work history line to adjust the base estimate for deviations from the base estimate situation in: workplace (KW), environment (K_E), materials (K_M) and task (K_T). This process is described below.

The schematic overview in Table A1.2 shows the available measurement data for the exposure assessment in this study. The majority of measurements were collected in the period between 1977 and 1986. Some data were available for the job title ‘loader’ in 1953 (derived from total hydrocarbon results) and service station attendant in 1969. For some job titles (checker/gauger, marine attendant, marine pumpman, and barrel washer), multiple IOL company industrial hygiene survey reports were referenced in the paper (Armstrong et al., 1996), which could not be easily tracked down during the site-visit. However, all industrial hygiene reports were indexed and searches by specific agents or jobs are therefore possible for future use. However, for this report we cannot be sure about the exact time period in which the measurements for these job titles were collected.

Table A1.2
Schematic overview of available measurement data

Job title	1950	1955	1960	1965	1970	1975	1980	1985	1990
Plantman, plant mechanic								Verma 1986	
Fuel delivery driver (agency)								Verma 1986	
Fuel delivery drivers (large terminal)						Saunders 1977			
Garage mechanic, pump mechanic						Tims 1977/1978			
Checker, gauger									
Barrel filler						Saunders 1978			
Loader	SODC 1953					Stacey 1977			
Marine deck attendant									
Marine pumpman									
Agent, acting agent							Coker 1984/1985	Verma 1986	
Air & service station attendants				Parkinson 1969				Cheng 1985/1986	
Barrel washer									
Facility background job							Conrad 1982		
General population background							Wallace 1981-1984		

When more than one source was used (see Table A1.2) it is unclear how measurement results from these different reports or papers were combined to construct the BE.

Gaps in knowledge were filled based on expert judgement from the industrial hygienists derived from several sources of information, which have been described before. This process was structured by the use of K-factors, but has not been described in detail.

Eleven out of 1086 (=1%) work history lines contained a job title, which did not fit one of the job titles in table A1.1 or table A1.2 (see Table A1.3). For those job titles, exposure to benzene could not be estimated based on the available exposure data from the selected reports and papers. Initially, exposure estimates were not assigned to these work history lines. Later in the process in the final dataset, these gaps in exposure estimates were filled based on a professional judgement of the industrial hygienist or based on previous and/or subsequent jobs (so called “bridging”).

Table A1.3
Job titles that could not be included in one of the job categories

Job title	Start	End
ASPHALT TECHNOLOGIST	1949	1949
ASST. SIGN PAINTER	1953	1953
BOILERMAKER	1919	1953
LABORER	1909	1948
OUT DOOR ADVERTISING	1923	1927
P. F. HELPER	1927	1933
PAINTER	1917	1919
	1927	1931
PAINTER'S HELPER	1919	1949
PROCESS FIREMAN	1936	1937
UTILITY MAN	1953	1953

All time periods for which no exposure measurements were available (see Table A1.2), had to be extrapolated backwards in time based on information sources, which are listed in section A1.7.2 by the use of exposure modifiers (K-factors) as described previously.

A1.8 ESTIMATION OF INDIVIDUAL EXPOSURES

As described above, exposure concentration estimates were made for all distinct lines of work history rather than by linkage to a JEM. Cumulative exposure (in ppm-years) was calculated by multiplying the workplace exposure (WE) by length of time in a job. In addition, exposures were lagged by 5, 10, and 15 years.

Gaps in work histories were filled by Tom Armstrong and Eileen Pearlman and reviewed and approved by an independent scientific advisory board based on previous and/or subsequent work history lines. For some job titles, that could originally not be estimated, a "professional judgement" was used as exposure estimate.

A1.9 STATISTICAL ANALYSES

Statistical analyses were carried out using the statistical packages SAS and EGRET, the latter being a specialised epidemiological package. Odds ratios were calculated, calculating odds ratios from comparisons of cases with their matched sets of controls. After exclusion of individuals with inadequate work histories, there were only 14 leukaemia cases and 55 controls for analysis.

A number of analyses were performed, calculating odds ratios with different grouping limits on the estimated cumulative exposures to benzene, with variously lagged exposures, and with duration of exposure as an alternative metric. In each case, odds ratios were calculated in comparison with the lowest exposure category. There were also attempts to fit conditional logistic regression models to allow adjustment for two potential confounders, smoking status and family history of cancer, and fit a continuous cumulative exposure. The latter produced an odds ratio of 5.53 for family history of cancer, suggesting a strong familial component of leukaemia risk. Models introducing smoking were unstable; this is not surprising, since smoking data were available for only seven of the cases, who were all smokers, so that logistic-based models including a non-smoker/smoker comparison would inevitably fail to converge.

The results of the odds-ratio analyses showed no evidence of a relationship of leukaemia risk with benzene exposure, no matter which set of group limits was used. It seems clear that the absence of a relationship was not an artefact of the positioning of group limits. Analyses comparing groups distinguished by their average or maximum benzene intensity (airborne concentration) or probability of dermal exposure also showed no relationship.

Logistic regression models showed no significant relationship with cumulative benzene exposure, nor with mean intensity, nor with duration of exposure. The significance level for duration was 12%, and the authors tried other models including the square of duration; this fitted better than linear duration, but both seemed implausible in that the relationship was inverse, with higher risks corresponding to shorter durations.

Given the small numbers of case-control sets, investigation of interactions was not attempted, and cross-validation of models within the data set was not feasible.

The authors reported power calculations indicating that, had the exposure group with cumulative exposure higher than 45 ppm-yrs had a 2-fold increase in leukaemia risk, their study would have had only a 16% chance of detecting this excess at a 5% significance level.

A1.10 VALIDATION AND RELIABILITY

Given that the sources of the data behind cohort definition were official written records, it is not likely that any external or alternative source could have been used to validate the data extracted. Indeed, the record sources are exactly of the type that might have been used to validate data collected from the memories of cases, controls or proxy respondents. The personnel systems used to define the cohort are likely to have been almost entirely complete and relatively error-free. The medical records, used to provide information on confounders and to verify occupational histories, would have been partially self-validating, in that separate records were made of details at multiple visits.

As an independent check on the exposure assessment method, 35 work history lines were constructed based on IOL survey reports with measurement data. These artificial work history lines were (blind to the exposure assessors) added to the list of work history lines as a validation process. All measured exposure levels were within estimated confidence intervals. The percent difference between estimated and measured benzene exposure for these 35 work history lines was 22.9% on average and ranged between 0% and 130%. No rank correlation between measured and estimated exposures was calculated.

The principal sensitivity analyses performed were those that experimented with different cutpoints and groupings of the cumulative benzene exposure variable, and these showed that the main conclusions were not sensitive to the choice of those groupings.

A1.11 CHECKING RANDOM SAMPLE FOR ERRORS OR INCONSISTENCIES

A random sample of cases and controls were checked for errors or inconsistencies. Two cases (ID numbers: 5144 and 10202) and four controls (ID numbers: 9762, 9702, 67369200915, 611136971222) were randomly selected from the case-control data set.

The exposure assessment was checked by using the Excel spreadsheet with exposure estimates, modifying factors, certainty score and dermal exposure score, which was used by industrial hygienists and compared those with the exposure estimates in the final SAS dataset, which was used for the statistical analyses in the nested case-control study. One (probably rounding) error was found for ID 67369200915, where the benzene workplace exposure estimate was found to be 0.353 ppm in the spreadsheet and 0.350 in the SAS dataset. This has not influenced the results of this study, because exposure estimates were rounded off at two decimal places in the analyses (in this case to 0.35 ppm).

This rounding off, however, did influence the work history lines that were assigned the population background exposure level of 0.005 ppm, because these were subsequently rounded off to 0.01 ppm, which is equal to the site background exposure level. This means that in the statistical analyses, site background exposure levels and population background exposure levels were treated the same. In total, 390 out of 1086 (=36%) work history lines were assigned a workplace exposure estimate of 0.005 ppm.

Because this study did not have a job dictionary, with information on which job titles were brought together in a job category (Table 1), we were not able to check the assignment of BEs to a certain job title for these six subjects.

APPENDIX A2: REPORT OF SITE VISIT TO UNIVERSITY OF BIRMINGHAM: IP STUDY

A2.1 INTRODUCTION

Since the mid 1970s the Institute of Petroleum (IP) (now the Energy Institute) has sponsored the follow-up of two large historical cohorts of workers in the oil industry. The cohorts consist of approximately 35,000 men from eight UK oil refineries, and 24,000 men from UK distribution centres.

The cohort studies were originally carried out by Michael Alderson and Lesley Rushton at the Institute of Cancer Research. The follow-up of the cohort studies and the case-control study nested in the distribution workers' cohort were carried out by Lesley Rushton, at Nottingham University. In 1995, Fintan Hurley and John Cherrie carried out a data quality audit of certain aspects of the study, focussing in particular on the reliability of work histories, the site histories, the assessment of historical exposure levels and the calculation of estimated cumulative exposures to benzene. They found that the work had been carried out to a generally high standard, but observed a number of areas where documentation was incomplete, and recommended *inter alia* that these omissions should be rectified.

In 1997, curatorship of the study was transferred by the sponsors to Tom Sorahan at the University of Birmingham. In preparation for the transfer, a great deal of effort was put into organising, collating and preparing both the study records and accompanying documentation.

During the visit for the present study, we saw several filing cabinets of documents, and many archive boxes containing study records (mostly cohort registration forms and death certificates). We also investigated the contents of a folder of PC computer files, designated 'Machine A' in the documentation, that had been taken from a PC physically transferred from Nottingham to Birmingham. We had been provided with documentation of the contents of this PC, and the contents were in agreement with the documentation. As the visit progressed, however, we noticed an absence of certain other classes of materials, including:

- detailed material on the outputs of the exposure assessment;
- reports that were used to construct exposure estimates;
- computer programmes or routines for combining data preparatory to statistical analyses;
- completed data files for statistical analyses;
- programmes for statistical analyses;
- outputs from statistical analyses.

It was clear that, without these computer materials, our review could be only partial, particularly in areas relating to the documentation of a number of key procedures in the exposure assessment and the statistical analysis. One lever-arch folder was discovered that listed and documented some data-collation programs written in the statistical system SAS, but we could find at that stage no corresponding computer files or hard copy of the programs or their output.

Later in the visit Eileen Pearlman (who had been seconded from ExxonMobil to assist us in understanding the work of the Exposure Assessment Task Force on this project) discovered, in one section of one of the filing cabinets labelled 'Machine B', an additional section of documentation and three data tapes of the sort (BC6250) commonly used for backup and archiving around 10-15 years ago.

The newly discovered documentation listed a large number of PC folders different from those we had seen on the Birmingham PC. These included SAS programs whose documentation we had already seen, data files and programs for statistical analyses in the epidemiological EGRET system. It was also clear from the numbering of the pages that this section of documentation had been intended to follow is sequence with that for Machine A. However, Tom Sorahan was certain that only one machine had been delivered to Birmingham. We examined that machine and found that it still contained the files described in the Machine A documentation, but not of the Machine B files.

From this, we were led to deduce

1. There were two PC computers at Nottingham, referred to in the transfer documentation as Machines A and B.
2. Machine A was transferred physically to Birmingham.
3. The three backup tapes were in the same drawer compartment as the Machine B documentation. Although they were labelled simply “Trolley Machine”, it seemed likely that they might contain the contents of the missing Machine B.

Neither Lesley Rushton nor Tom Sorahan has been able to confirm this scenario from documentation, but it seems at least plausible (and has now been confirmed to Lesley Rushton by the computer manager who is still at Nottingham). Birmingham sent the three tapes to a specialist data recovery bureau, but they reported that the data were in an unknown binary format.

Lesley Rushton then contacted IT specialists at Nottingham, who recognised the tape format, and produced a machine that would read the tapes found. When this was done, it was found that the three tapes involved did not form a complete backup set, so that, while some files could be extracted from them, it would not be possible to be certain whether any files were missing. Nottingham staff then uncovered a further supply of similar back-up tapes, and agreed to search these to see whether a complete backup might be identified. At the time of writing, we are informed that it has been possible to identify a large number of files from the IP study. We do not at present know whether the files found complete the inventory of what was supposed to be transferred with Machine B, or whether they represent only part of that catalogue. Because they were discovered so late in the investigation, we have not attempted to make any use of them, and in fact have not received copies of any of them.

We have been made aware that, prior to leaving Nottingham University, Lesley Rushton sent a copy of the IP study data to R Schnatter. On inspection, this is not a single file, but a database with a great many linked tables. It is possible that some of these tables may duplicate missing files from Machine B.

We strongly recommend that the process of cataloguing the extant files from the IP study should be carried to completion. This should involve:

1. Checking files recovered from the Nottingham tapes against the catalogue listing what should have been transferred to Birmingham as the contents of Machine B

2. Checking to what extent the contents of the database files held by Rob Schnatter provide material duplicating or additional to the Machine B files.

A2.2 COHORT DEFINITION

The original cohort was defined as all men who had worked in a number of UK distribution centres within three companies; this became four after 1976, when two components within one of the three became same separate companies. Eligibility required at least one continuous year’s service between 1 January 1950 and 31 December 1975.

Cohort members were identified from personnel and pension departments at company head and regional offices. Since membership was based directly on existing employment records, the cohort is believed complete.

For all cohort members, full name, date of birth, address, distribution centre, date of joining and last or present job were extracted. For leavers, date of leaving, reason for leaving (retired, died or other) and National Insurance (NI) number, where available, were recorded. In some cases, National Health Service number was available.

Manuscript records of the individuals in the cohort study are stored in a series of archive boxes at Birmingham University. They consist of standard forms with hand-written entries. The forms are two-part, with the name and address on one part and the occupational information on the other. The parts are linked by a unique identifier consisting of a letter identifying the company and main site, a two-letter code for the distribution centre, and a four-digit id number. In fact, the id numbers are unique within the initial letter, and the forms are stored by id number within the initial letter.

After processing, the forms were split in order to anonymise the cohort members. They are now stored in separate batches. For the cohort studies, no attempt was made to collect full occupational histories.

The cohort data forms have all been scanned, and their images are held on data CDs at Birmingham University.

The cohort study (Rushton, 1993) suggested an increase in leukaemia risks among distribution centre workers whose last recorded occupation was as a driver; this was at least part of the justification for initiating a case-control study.

A2.3 CASE IDENTIFICATION

Cases were men from the cohort who (a) died before 1 January 1993 with a mention of leukaemia (either underlying or contributory cause) on the death certificate, or (b) had a cancer registration of leukaemia. The definition of leukaemia was standard, as ICD-9 codes 204-208 inclusive (or equivalent in earlier ICD revisions)

Mortality tracing for occupational studies in the UK use as their principal source of information the National Health Service Central Registers, because almost all UK citizens are registered with the National Health Service (and attached to a GP practice). However, the unique NHS number indexing that register is seldom known to the individual, and rarely recorded in occupational records, so that tracing in the NHSCRs requires matching on full name and date of birth, with address sometimes being used to separate multiple matching. Typically, vital status can be ascertained in 95-97% of cases, with the remainder returned as emigrated or untraced. In the past, emigration has been the largest single cause of loss to follow-up.

In past times, it was also possible to search in the National Insurance system. In the UK cohort study, those untraced in the NHSCR were cross-checked in the NI system, and many found to be alive (either paying contributions or receiving benefits). This reduced the untraced rate to less than 1.5% in the 1989 follow-up, a very good result.

All qualifying cases from the cohort were chosen. There were 88 leukaemia cases identified from death certificates, and another three from cancer registrations alone. Given the high success rate of tracing for deaths, it is expected that the probability of having missed mortality cases from leukaemia would be small. For the cancer registrations, however, we note that cancer registration is not compulsory in the UK, and while it is believed that registration is now relatively complete, it is possible that this was not so in the years after 1971, when registration was introduced..

Case subtypes were allocated initially from the death or cancer registrations. Of the 40 cases identified in both systems, 9 had differences in the diagnosis, with the cancer registry usually being less specific. The more specific description was taken for analyses grouped by cell type. Attempts were made to verify the diagnoses from histopathology departments, and by publication date confirmation had been achieved for 38 of the cases. Of the remaining 53 for whom a reply was awaited or no histology record available, 3 had the diagnosis noted in company records, and 13 had the same diagnosis on death and cancer registry certifications. Of the 9 where these records differed, histology reports received for 6 all confirmed the more detailed diagnosis. None of the histology reports for other cases resulted in a change of diagnosis for the study.

The end date for cases was taken as the date of diagnosis where this was known, or the date of death otherwise.

Since the inclusion of cases did not depend on case co-operation, all identified cases were included in the data-gathering exercise.

A2.4 CONTROL IDENTIFICATION

Controls were selected without replacement at a target ratio of 4:1, from the same company, and a year of birth within three years of the case date of birth. Controls had to be alive at the time of the case's occurrence.

Since the inclusion of cases did not depend on case co-operation, all identified controls were included in the data-gathering exercise. It was subsequently discovered that two controls selected were not actually under follow-up at the date of the case occurrence, and eight of the controls for four cases (all from one company) had been incorrectly matched for age. All these controls were excluded from the analyses, plus one case that then had no valid controls. The algorithm for selecting controls we rechecked and found correct, so it is not clear how the error occurred, and no evidence that it had occurred in the other three companies.

A2.5 DATA COLLECTION IN CASES AND CONTROLS

The cohort data included only company and distribution centre,, date of birth, dates of joining and leaving, reason for leaving, and last known job. It was intended to collect as full as possible an occupational history for each subject, entirely from company records, either from personnel systems or from pension records. There was no use of proxy sources.

Data on smoking habits were collected, where available, from archived health records. Where data on previous employments were available in personnel systems, these were also extracted.

Data on smoking habits were available for only a small minority, and smoking habits were recorded as "unknown" for 79 cases (88%) and 312 controls (88%)

A2.6 WORK HISTORIES IN CASES AND CONTROLS

An attempt was made to collect work histories in as much detail as possible, to cover each job held within the company, with start and end dates and site.

The major source of information on work histories was personnel records, but availability of these varied according to the record retention policies of the four companies involved. Company A had retained all of its personnel records, but companies B and C retained records for only seven years after an employee left. Company D retained many of their personnel records on microfiche and microfilm. For those study members for whom a personnel record could not be found or was incomplete, the information was obtained from a variety of sources including pension records, medical records, and interviews with retired or long service staff.

Staff extracting the work histories were blind as to the status, case or control, of the subjects whose data they sought. The data were recorded and updated on standard pro-formas, and these are stored alphabetically by surname within company (with companies A and D combined) in filing cabinets at Birmingham University. For company C, the file on each subject contains copies of the personnel records etc. used to compile the work histories. For company B, the work histories had been compiled at the company offices, and copies had not been taken of the source documents. For companies A and D, copes of the source documents had been produced (and were seen with the pro-formas in the files by Fintan Hurley and John Cherrie in 1995). They are not, however, now in these files. Lesley Rushton remembers that the Chief Medical Officer of the Company concerned requested that these should be returned to the Company once the study had finished and this was duly carried out. It is not known what happened to the work histories then, nor whether they were destroyed by the Company.

This means that the only company for which the source documents for the pro-formas could be consulted was company C. We did not attempt to interrogate the microfilm sources. Without these, it appears that at present it is not possible to check the process by which the work history pro-formas were compiled, except for the small number of subjects from company C.

The job were not coded as such, but it was recognised that the titles used in company records did not always reflect very closely the tasks that the person actually carried out, and job titles differed between companies. A ‘job-task dictionary’ was developed, using information from company and union records, supplemented by a large number of interviews. The job-task dictionary gave a description of the tasks performed under each job, from both official records and anecdotal information.

Where work histories were incomplete, assumptions were made, including:

1. use of mid month or mid year if days and/or months were unknown;
2. if there were gaps in the dates for either a job or a location, taking a date midway between known dates;
3. where a job history was largely unavailable, using a ‘typical’ work history for that job. For example, for a man who started work aged 16, worked for over 30 years, with ‘driver’ as his last job title (from the cohort records), it would be assumed that he started as a lorry boy, became an adult mate or driver’s mate, and a driver at age 21, the age at which a heavy goods licence could be issued. In the absence of any further information, it would be assumed that the entire career was at the location indicated as the final location.

Company A work history records were mainly complete, and demonstrated that men with last jobs as clerical workers, operators or drivers typically had spent most of their employment in this type of job, lending some confidence to the imputation of typical job histories for such jobs. However, those with managerial or supervisory last job titles could have a variety of different work histories, and movements between sites, so imputation for these would be less reliable.

As a measure of the quality of the work history, each study member was assigned a ‘job confidence’ code of 1, 2 or 3, where

1. indicated a complete work history, or one for which only the last job was known but for which the length of employment was under 10 years.

2. indicated a partially complete work history for which an assumption as described above had been made
3. indicated a poor work history, where only the last job title was known, this was a supervisory or managerial post, and the man had worked for more than 10 years.

There were no cases and 8 controls labelled as job confidence code 3; 23 cases and 68 controls in code 2; and 67 cases and 278 controls in code 1.

A2.7 EXPOSURE ASSESSMENT AND ASSIGNMENT

A2.7.1 Background

Approximately 315 terminals were identified from the work histories, over 95% of which were closed by the time of data collection. All the remaining (operating) company-owned distribution terminals involved in the study were visited and terminal characterisation proformas (developed with the company occupational hygienist) completed. The proforma developed as the study progressed to focus on key areas of interest. Whenever possible, interviewees were also asked about closed distribution terminals that were pertinent to the study. Telephone interviews with retirees were also conducted to obtain information on closed locations. Other sources of information included site plans, booklets, photographs, and company magazines, and other material available from company libraries, property and engineering departments.

Because the case-control study was focussed on low levels of benzene exposure, exposure measurement results didn’t come close to the OEL for benzene. Mostly, measurements were taken routinely and not specifically focussed on marketing and distribution workers with low levels of benzene exposure.

A2.7.2 Sources that were used for compiling historical data about the plants:

- Site visits
- Company proformas
- (Telephone) interviews with retirees
- Site plans
- Booklets
- Photographs
- Company magazines

All these different sources of information were stored per company and location for use in the process of extrapolating measurement data to different sites and time periods.

A2.7.3 Exposure measurements

Over 200 occupational hygiene reports and articles relating to petroleum distribution jobs were collected from the literature, participating companies and industry association sources. Each participating company report was given a unique identifier and the data tabulated. All personal exposure measurements were scrutinised to establish whether they were sufficiently well characterised with regard to the job titles, the benzene content of the fuel handled, the range of fuels handled, the product and air temperatures and the type of technology in use at the specific locations. Insufficient characterisation of the studies meant that less than 20% of the reports collected were used for BE derivation.

There were few full shift benzene monitoring data available for ‘terminal operators’. Therefore, the separate tasks undertaken by terminal operators were reviewed. Where no or few data existed for the separate tasks, additional monitoring was carried out for this case-control study. Based on those monitoring results, the base estimate for ‘terminal operators’ was divided into five separate time periods (1910-25, 1925-45, 1945-60, 1960-75, 1975-92) and terminals were classified into three sizes (small, medium, and large), primarily based on the number of drivers employed.

During the site visit in Birmingham, most of the occupational hygiene exposure measurement reports on which the base estimates were based, appeared to be missing. For some bases estimates the individual measurement data points were listed without any reference to a report. For some BEs the measurement data points referred to a report that was not available in Birmingham during the site visit. And for some BEs there were no data at all to discover where the BE was derived from. However, during the Fintan Hurley and John Cherrie audit (Hurley et al., 1995) carried out in 1995 in Nottingham, these exposure measurement reports were all available in Nottingham as well as the computer spreadsheet files with the exposure measurement results that were used for base estimate derivation. At that time Hurley and Cherrie concluded that the derivation of base estimates had been carried out in a careful manner and no data entry errors were found. However, they recommended that all reports collected as potential data sources for the IP project should be catalogued and copies should be archived (Hurley et al., 1995), of which no evidence was found in Birmingham in 2004. A possible reason for the missing reports is that company reports were returned to the respective companies at completion of the study.

Base estimates for ‘population background’, ‘jetty operator’, and ‘drumming’ were derived from CONCAWE reports (report no. 4/87 and 1/94) and for the base estimate for ‘top submerged driver’ company industrial hygiene reports were available in Birmingham.

Because most of the company industrial hygiene reports were not available in Birmingham, we were not able to determine why measurements were taken, what kind of measurements were taken and what background information was collected for the measurements.

All data that were used for base estimate derivation were statistically tested using the statistical program REST (developed by one of the companies). During the site visit, none of these (computer or hard copy) data were available in Birmingham.

Arithmetic mean exposure levels were calculated from the reports (most of which were not available for this review) per job title and used as Base Estimates (BE) for that job title (see Table 1). When necessary, the exposure data were standardised to 3% benzene content in the petroleum, the handling of 100% petroleum, and a temperature of 10°C, to enable data from different sources to be combined. Subsequently, those Base Estimates were adjusted for factors that may have influenced exposure to benzene. These factors were extensively reviewed and discussed at an international workshop held at the University of Nottingham in 1993 and divided into six modifying factors (‘K-factors’) for:

- job activity (JA), to adjust for known differences in the job title to be estimated form those of the BE;
- number of loads (JL);
- percentage of benzene in the gasoline (PB);
- mixture of products (PM) to adjust for other products such as kerosene or diesel;
- air temperature (MT);
- loading technology (TS), particularly to account for differences between the top splash and top submerged loading techniques.

The retrospective workplace exposure estimate (WE) for each work history line was obtained by adjusting the base estimate (BE) for these modifying factors (K-factors): $WE = BE \times K_{JA} \times K_{JL} \times K_{PB} \times K_{PM} \times K_{MT} \times K_{TS}$

During the site visit, there was no data file available to indicate which BE and K-factors were assigned to the work history lines. A hardcopy printout of such a data file was found in the documentation, but we cannot be sure that this was the final version.

The statistical software REST was used to test the data for (log)normality and to determine the variation in the data points (of which no output was available during the site visit).

For the assignment of BEs and K-factors to work history lines, no certainty score was used to indicate the reliability of the assignment, as was done for the IOL study.

Because most of the measurement data were not available during the site visit, we cannot be certain whether the measurements were collected for the cohort or that they came from outside the cohort. Nevertheless, we assume that most measurements were collected within the cohort (company industrial hygiene reports) and that only few data came from outside the cohort (CONCAWE reports).

A2.7.4 Assigning exposures to a Job-Exposure Matrix (JEM)

Base estimates were calculated for inhalation exposure only. In addition, a ranked grade of dermal contact was assigned to each work history line (none, low, medium, or high) by the exposure assessment task force (EATF). No biomarker data were used.

The base estimate that best suited to the job undertaken by that individual was assigned to each work history line. The list of job titles is reported in table 3 in the exposure assessment paper (Lewis et al., 1997). A copy of this table is presented in Table A2.1.

Subsequently adjustment factors (K-factors) were assigned to each work history line to adjust the base estimate for deviations from the base estimate situation in: job activity (K_{JA}), number of loads (K_{JL}), percentage benzene (K_{PB}), product mix (K_{PM}), air temperature (K_{MT}) and loading technology (K_{TS}). This process is described in section A2.7.3. Modifying factors were assigned after a review of available information, including the individual terminal characterisation files.

Because the job title ‘terminal operator’ was such a broad job category with several task activities, additional monitoring data were collected to make a distinction for ‘terminal operator’ between terminal size and time period.

The full work histories were not all available at the beginning of the estimation process. Accordingly, estimating exposure in this nested case-control study was an iterative process, in which the collection of work history data and information about different companies and sites was done at the same time as the base estimates were assigned. Therefore, several preliminary, incomplete printouts of base estimates per work history line were generated, checked and completed in the course of the study.

A2.7.5 Base Estimate information

As shown in Table A2.1, 75% of all work history lines were assigned the BE for ‘site background’ (41%) or ‘driver top submerged loading’ (34%).

Few computer files containing work history lines and base estimates were available to this review. However, these files did not match in number of work history lines. In addition, base estimates and modifying factors were coded and no analysis files were available to this review. Therefore, we cannot be sure about the correctness of the data files and we cannot comment on gaps in knowledge for assigning base estimates and modifying factors to work history lines. Furthermore, we do not know where the exposure measurements, which were used to construct base estimates, came from. Therefore, we do not know how much extrapolation was required in this study.

Table A2.1
Base estimate benzene concentrations by job

Job title	Base Estimate concentration Benzene (ppm)	Frequency of occurrence in the work history lines of the study
Site background	0.016	660
Small terminal, operator 1910-25	0.88	1
Small terminal, operator 1925-45	0.57	9
Small terminal, operator 1945-60	0.16	40
Small terminal, operator 1960-75	0.18	17
Small terminal, operator 1975-93	0.28	0
Medium terminal, operator 1910-25	2.50	0
Medium terminal, operator 1925-45	2.36	11
Medium terminal, operator 1945-60	2.32	29
Medium terminal, operator 1960-75	0.23	31
Medium terminal, operator 1975-93	0.16	11
Large terminal, operator 1910-25	2.93	0
Large terminal, operator 1925-45	3.49	4
Large terminal, operator 1945-60	3.22	12
Large terminal, operator 1960-75	0.68	21
Large terminal, operator 1975-93	0.23	13
Benzole terminal, operator 1925-45	7.34	6
Benzole terminal, operator 1945-60	5.00	7
Driver without loading	0.13	66
Driver top submerged loading	0.40	549
Driver bottom loading	0.26	2
Vehicle maintenance	0.16	15
Population background	0.003	26
Airfield operator	0.11	34
Jetty operator	1.38	14
Service station attendant	0.11	4
Drum filler	8.20	9
Rail car loader	1.27	1
Pump and tank fitter	0.48	20
Laboratory	0.19	4
	Total	**1616**

What we do know is that in the event that dates, job title, or job location data were missing or no site characterisation data were available, the exposure assessment team discussed the assignment of specific values, reaching a consensus decision and providing a recorded comment as to the choice of modifying factors made.

A2.8 ESTIMATION OF INDIVIDUAL EXPOSURES

Cumulative exposure (in ppm-years) was calculated by multiplying the workplace exposure (WE) by length of time in a job. Furthermore, each work history line was given a code, which characterised the type of peak exposure experienced for that job at that time and site, according to the duration (1-15 min. or 15-60 min.), frequency (daily, weekly, monthly) and intensity (1-3 ppm or >3 ppm) of the peak.

Because data files on the assignment of base estimates and modifying factors to work history lines were missing, we cannot determine the number of missing data and the extent to which imputation was required. In addition, it was unclear where the exposure measurements, which were used to construct base estimates, came from, which made it impossible to measure the extent to which extrapolation was needed in this study.

A2.9 STATISTICAL ANALYSES

All data were organised in a Paradox database. Data manipulation for the preparation of files for statistical analysis, and descriptive and exploratory analyses, were carried out using SAS. The epidemiological package EGRET was used for case-control analysis. As described in the Introduction, we believe that many files from the data manipulation and analyses were originally on Machine "B", and have not been provided at the date of writing, so it has not been possible to review these.

The analyses were carried out using standard statistical techniques for matched case-control studies, i.e. conditional logistic regressions categories in different ways and as continuous predictors.

For these analyses, cumulative exposure (ppm years) was categorised in quintiles and in four pre-defined groups. Analyses were also done on categories of average intensity of exposure and of duration of employment. Indicators of skin contact and likelihood of peak exposure were also included.

Potential confounders or effect modifiers included in the analyses were smoking, employment status at end date, predominant socio-economic class, age and date started work, ever had a previous job, and ever a driver in a previous job.

Analyses were repeated omitting exposures less than five or ten years before the diagnosis date, to allow for latency.

A wide range of models were fitted, and analyses for subtype of leukaemia were also carried out.

The combined analysis for all types of leukaemia was based on 90 cases and 354 controls. The results showed no trend across cumulative exposure categories, and the coefficient for exposure as a continuous variable was very close to 1.0. There was also no relationship with average intensity of exposure, but a consistent doubling of risk in those employed in the industry for more than 10 years. Neither acute nor chronic lymphoblastic leukaemia showed a relation with cumulative exposure. However, risks for acute myeloid and monocytic leukaemia appeared to show a relationship with and OR of 2.8 for the highest exposures against the lowest group, and a similar range of OR across categories of intensity of exposure. However, models fitting these variables as continuous predictors did not show statistical significance. Chronic myeloid leukaemia showed no relationship with any of the exposure variables.

No explicit power calculations were performed. However, this was by far the largest and therefore the most powerful of the three case-control studies with explicit estimates of benzene exposure [we don't know how good the estimates were, because of missing data], and the numbers meant that the analyses of some subtypes were based on reasonable numbers.

A2.10 VALIDATION AND RELIABILITY

The study attempted to use the best quality data for all the variables. All identification data and work histories came from documentary sources, and there is no source of data against which they could be cross-validated. Where these existed, it is unlikely that they are incorrect, but it is known that they are incomplete for a proportion of the subjects.

An evaluation of the representativeness of selected BEs was carried out by comparing them with data from sources not used in this study: European studies (CONCAWE report no. 7/94), other UK data (Parkinson et al., 1971), and Canadian distribution workers (Armstrong et al., 1996).

A number of sensitivity analyses were carried out, particularly analysis omitting individuals whose work histories were less reliable. It is reported that these showed similar patterns in general; but that the reduced numbers showed no increase in risk with exposure for acute myeloid and monocytic leukaemia. In the analyses with omission of recent exposures for latency, there was a tendency for ORs to increase for acute myeloid and monocytic leukaemia, and to reduce for chronic lymphocytic and chronic myeloid leukaemia.

It was also noted that some of the models did not fit well to the data, as judged by overall goodness-of-fit tests.

A2.11 CHECKING RANDOM SAMPLE FOR ERRORS OR INCONSISTENCIES

From the data files, we selected 30 subjects for whom to audit and check the data handling. As noted above, the source documents from which the work histories were extracted have not been retained in the files, so it was not possible to audit this step. The summary work history sheets have been retained, in many cases with multiple generations and extensive hand annotations. We checked that the work history lines in the data files matched the hard copy summary of the work history. This was made harder because the data were held in a database format for which we did not have the matching software. We discovered that the database files could be opened using the IOM's copy of the statistical package Genstat. It turned out that Genstat coded and stored data on dates differently from the original package, but we were able to effect a simple calculation to make the translation.

The results showed that in the large majority of cases there was agreement between the hard copies and what was in the files. In a few cases it was apparent that there were more lines in the data files than on the proforma, but we were satisfied that this arose because a line had been split during the exposure assessment. There were also a few small discrepancies in dates. However, we found no serious discrepancies, and noted that where dates were recorded on the proforma as assumed, they were correctly labelled as such in the computer files. We discovered one case where there was a work history in the computer files but no work details in the proforma. It was presumed that the computer work history, which described the man as a watchman, location Plymouth, had been constructed as described earlier, either from interviews or based on assumptions extrapolating from the mlast job held. It was not possible to verify this, which suggests that the process of filling in work histories did not always include the copying back of this information to the data collection proformas. This would be in line with a view that the computer data files were the master version.

APPENDIX A3: REPORT OF SITE VISIT TO ADELAIDE AND MONASH UNIVERSITIES: HW STUDY

A3.1 INTRODUCTION

The Health Watch cohort study of Australian petroleum industry workers, based on periodic health surveys of workers, began at the University of Melbourne in 1980. A nested case-control study was begun in 1988, with quantitative exposure assessment beginning in 1995. In 1999, the two studies were transferred. The parent cohort study was placed at the University of Adelaide, under Dr Richie Gun, and the nested case-control study was placed with Dr Debbie Glass, at Monash University, Melbourne. The latter allocation provided some continuity, since Debbie Glass had worked on exposure assessment for the case-control while at the University of Melbourne.

Site visits were therefore made to the University of Adelaide (by BGM alone) and Monash University (by BGM & WF).

A3.2 COHORT DEFINITION

The Health Watch studies were sponsored by the Australian Institute of Petroleum (HW) and were intended to include all industrial workers from participating companies within the Australian petroleum industry. Participation was agreed with all companies, with the possible exception of a very few very small enterprises. The Castrol company withdrew from the HW in 1994, but was included in the earlier surveys so Castrol employees have been included in the case control study.

The surveys were intended to cover all petroleum industry workers in refineries, onshore and offshore production facilities, and the distribution sector, including airports and terminals. City centre offices and all sites employing fewer than 10 persons were to be excluded. The surveys were to be restricted to blue collar workers, but in fact some managerial staff also participated.

Participation was voluntary, and the study was introduced to workers at on-site presentations. Target populations were defined from company employment rolls, which provided information about joining and leaving dates plus some details of jobs held. Participation in the surveys was generally good, of the order of 90%.

The initial recruitments were by face-to-face interviews in four surveys at approximately 5 year intervals between 1980 and 2000, although this was later changed to a system of postal questionnaires. As the surveys progressed, new potential recruits and important events such as leaving, retirement and death were self reported and identified from company rolls initially every 6 months, later annually. These regular updates also allowed recruitment of eligible members missed at previous surveys. As a result, it is believed that the coverage overall of the eligible population may have been nearer to 95%. The cohort is now closed to further recruitment.

The interviews were informed by the company data on start and finish dates. A consent form, allowing extraction of relevant data from the company records was signed by the participant. A detailed work history was taken from the worker, and smoking habits, alcohol consumption and any recent health problems were also recorded. In the first rounds of surveys, the job history was intended to cover only the previous five years. A complete job history was sought in the 3rd survey, and in the early 1990s attempts were made to fill in missing early years, by corresponding with the study subjects who had retired or left the industry so they did not take part in the 3rd survey.(see section A5)

The hard copy records of the Health Watch records are stored in off-site archives, but the records for all cases and controls selected for the nested case-control study are stored on-site. The survey records held by Monash University for cases and controls are direct photocopies of these.

The University of Adelaide maintains the data in a Microsoft Access database that was inherited from Melbourne University. According to Adelaide staff, they do not have (and have never had) complete documentation for this database in its entirety. While they have documentation for the principal variables, including all those used in actual analyses, there are others that are not fully documented, and whose exact meaning is not clear today. This is presumed to relate primarily to intermediate variables created during the early processing of data, but it implies that many stages of that data processing cannot now be traced in their entirety. This is not considered a major concern for the cohort study, since there is evidence from the early Health Watch reports that data checking during the cohort construction was extensive.

The Health Watch surveys have analysed mortality and cancer registration rates for a number of causes, in comparison with national rates. These analyses identified, for all death causes combined, a "healthy worker effect", as would be expected in an employed cohort; a deficit of lung cancer; rates close to expectation for Non-Hodgkin Lymphoma; and a suggestion (not statistically significant) of raised rates for multiple myeloma and for all leukaemias. The latter result was difficult to interpret, since when broken down by type, none of the four main types showed an excess, but there was a small excess in "other leukaemias". These analyses were based on only 27 incident cases of leukaemias (or 16 deaths), and therefore subdivision of types leads to very small numbers. However, an earlier report from the study had recorded higher SIRs, including one reported as significant for myeloid leukaemia (ICD-9 code 205), based on 15 cases (Health Watch Tenth Report, 1998).

A3.3 CASE DEFINITION

The case-control study was set up because of general interest in benzene exposures, and in the light of the findings of the cohort analyses. Cases were defined as "men in the Health Watch cohort who reported a newly diagnosed lympho-haematopoietic cancer to Health Watch (either by himself or by his family) that was confirmed by pathology report, cancer registration, letter from a medical practitioner, or death certificate". The requirement for self-reporting was an ethical requirement; without it, the study team could not pursue efforts to confirm cases. Cases found in the cancer registries who had not self-reported could therefore be included only if lost to follow-up or known to be dead.

The identification of cases was part of the process of identifying more generally disease incidence and mortality within Health Watch. Because the diseases of interest were cancers, they could be traced in the cancer registries. Cancer registration is compulsory in Australia, and tracing initially took place in individual state registries, and latterly in the national registry (National Cancer Statistics Clearing House – NCSCH) established during the life of the study, which collates data from all the states. Data were also cross-checked with the National Death Index (NDI).

Seventy-nine cases selected for the case-control study met the definition of lympho-haematopoietic cancer, Thirty-three of these were leukaemias. These numbers are slightly higher than in the published paper on the cohort study, because the case-control analysis happened later. Casehood was confirmed as described above, and for cases where diagnosis was uncertain the medical information was reviewed by a haematologist, who confirmed or revised the cell type involved. One other man found in the cancer registry had not self-reported, and so was not included.

There appear to be no published papers assessing the reliability or completeness of the Australian cancer registration system. The systems were set up with procedures as specified by IARC, who have reviewed the procedures and published the data collected (Parkin *et al*, 2002). Anecdotally a great deal of effort goes into ensuring completeness, and good liaison over the contributions from the state registries. Richie Gun did a study comparing and cross-validating registrations in NCSCH with self-notifications. He found 2.5% of cases were in state registers but had not propagated to NHSCH (although this may have been partly because some were still in the process of notification), and another 2.5% had not entered the system. For deaths, a number came up that were not in the NDI, but mostly because the deaths had happened abroad. It therefore seems unlikely that the discovery of cases could have been seriously deficient.

A3.4 CONTROL IDENTIFICATION

Controls were selected from the database of the Health Watch cohort study. Five controls were selected for each case, matched on year of birth and eligible (i.e. not a case) at the time of the case's diagnosis. Sampling was with replacement, with the result that 5 workers were used as controls more than once.

Almost all of the cases were identified, and their controls selected, while the study was held at Melbourne University. Only three new leukaemia cases were added to the case series after the cohort study was transferred to Adelaide.

A3.5 DATA COLLECTION ON CASES AND CONTROLS

All the data on individual subjects, both cases and controls, came from the Health Watch records. These included work histories, taken from company records cross-validated with self-reports from the surveys. Where these conflicted, the more detailed source (usually the self-report) took precedence.

The Health Watch records also provided data on self assessed smoking habits, on typical alcohol consumption and on previous employments as potential confounders. Smoking habits were recorded as typical daily separate consumptions of cigarettes, pipe and cigars, age started, and notes of whether consumption had changed markedly in the past. For ex-smokers, age started and stopped plus typical daily consumption of cigarettes (only) were recorded. Alcohol consumption was recorded as typical number of days when a drink was taken, and usual average number on those days.

The case control study did not collect any new data from identified cases and selected controls, and there was no use of proxy sources in the collection of these data.

A3.6 WORK HISTORIES IN CASES AND CONTROLS

As already described, work history data were collected from the individual Health Watch subjects during the health surveys. The first round, which recruited the large majority of the subjects, took place between 1981 and 1983. The work histories taken then, by face-to-face interview, were intended to cover the previous 5 years' employment, and therefore could have missed considerable periods of early employment for employees of long standing. The second round (1986-87) used the same procedures, but most of those included would have been recruits since the first round, and therefore the five-year work history taken would have been complete for most of these. At the time of the third round (1991), complete work histories were taken and efforts were made to fill known gaps for all participating study subjects and an attempt was made to complete job histories for all members who had left the industry before third survey. At the fourth round, in 1994, histories were updated for participants, Work histories included information on location, dates and broad job classification, but did not describe specific tasks carried out.

This still left some work histories incomplete, but a check at University of Melbourne for the first 65 case-control sets, including 390 subjects, showed only 10 subjects that had incomplete work histories; for some of these, work histories were augmented from company records. The proportion of incomplete work histories remaining is thus low, of the order of 2%. It is not known whether these can be identified in the final data files, but it is unlikely that their inclusion or exclusion could affect results or conclusions.

Copies of all the survey records for the cases and controls are stored at Monash as case-control sets, i.e. one case and all the corresponding controls. We chose a representative sample of three of the leukaemia case sets, chosen to cover the earliest, middle and latest recruited cases. We were provided with anonymised photocopies of all the records in these sets, with all identifiers except the case series number and study subject number obliterated.

A3.7 EXPOSURE ASSESSMENT AND ASSIGNMENT

A3.7.1 Background

Historical changes in the plants were mainly described as important changes in the process technology or process conditions that could have changed the levels of exposure to benzene. Each site where subjects (cases and controls) were or had been was asked to complete a brief site assessment form (brief information about the activities, technology and products handled. Each site was then followed-up with a visit or telephone call. A questionnaire (based on the questionnaire devised for the IP study) was prepared to seek information on the history of the site, major changes in staffing, plant, product and technology. A report was prepared by a consultant summarising the available information on the benzene content of gasoline produced in each Australian refinery over the years. Information about jobs, activities, tasks and sites came from interviews with employees and ex-employees (including retirees) using job-specific questionnaires designed to seek information on the tasks, technology and products for relevant job activities.

A3.7.2 Sources used for compiling historical data about the plants

Sources that were used for compiling historical data about the plants included:

- Site assessment form, followed-up by site visit or telephone call
- Site questionnaire
- Report summarising the benzene content of gasoline in each Australian refinery
- Interviews using job specific questionnaires with:
    - Current employees, ex-employees (including retirees)
    - Readers of company and HW newsletter
    - Health Watch cohort members
    - Other individuals identified by personal recommendation

A3.7.3 Exposure measurements

Data for the base estimates was obtained primarily from the participating petroleum companies in Australia. For tasks where there were little or no Australian measured data, estimates were based on data taken from the literature, principally from CONCAWE documents (CONCAWE 1986, 1994b, 1997). Other sources used were Kawai et al. (1991), Kearney & Dunham (1986), Kramer (1989), Moen et al. (1995), Nordlinder & Ramnäs (1987), and Runion (1988). Some sources were added later to this list by the HW study team after the site visit and were therefore not reviewed in detail: CONCAWE 1981, 1987;

Rappaport et al (1987), Lewis et al (1997), Laitenen et al (1994), Foo(1991), Berlin (1988), Bader et al (1994), Popp et al (1994).

Task	Number of measurements	Reference
Driving and Unloading:	n=13, 15	Kawai et al (1991)
	n=7 short term samples	CONCAWE 97/52
	n=28	CONCAWE 4/87 & 97/52
	n=47	Rappaport et al (1987)
	n=7	Lewis et al (1997)
Dewatering	n=5, 4 short term samples	CONCAWE 7/94
	n=5	CONCAWE 97/52
Mechanic	n=?	CONCAWE 2/81
	n=4	Kearney & Dunham (1986)
	n>100	Nordliner& Ramnas (1987)
	n=8	Lewis et al (1997)
	n=22	Laitenen et al (1994)
	n=54	Foo (1991)
	?	Berlin (1988)
	n=5	Bader et al. (1994)
	n=20	Popp et al. (1994)
	n=13	Runion (1988)
Sampling	n=4, 4, 7	CONCAWE 7/94
	n=8,10	CONCAWE 97/52
Ship dipping and Gauging	n=2	Moen et al. (1995)
	n=8	Nordliner& Ramnas (1987)

Only personal exposure data were used and care was taken to ensure that the data used was a concentration for the actual period measured rather than an exposure normalized to an 8-hour average.

Refinery and terminal background values were taken from exposures measured on employees carrying out jobs that were considered by the petroleum industry occupational hygienists to have been non-exposed. Exposure monitoring data from the New South Wales and Victorian Environmental Protection Authorities was used for the urban and rural background values (Nelson, 1995; Wadge, 1996).

The reason why measurements were taken was:

1. *Participating petroleum companies in Australia:* measurements were routinely taken.
2. *CONCAWE report no. 3/86:* summary of data on exposure to benzene during the manufacture and distribution of gasoline in Europe.
3. *CONCAWE report no. 7/94:* summary of data on exposure to benzene during the manufacture and distribution of gasoline in Europe.
4. *CONCAWE report no. 97/52:* The report provides information on estimated release rate; measured human exposure data for occupationally exposed groups, consumers and the general public; and environmental exposure data for air, water, soil and sediments.
5. *Durand et al., 1995:* to evaluate the effectiveness of respirator selection during the cleaning of oil separators.

6. *Kawai et al., 1991:* to examine the extent of occupational exposure of tank truck drivers to benzene and other aromatic solvents.
7. *Kearney & Dunham, 1986:* to describe a sampling technique and analytical method to determine levels of ambient gasoline vapor in low concentration areas.
8. *Kramer, 1989:* to quantify short- and long-term exposure of underground storage tank contractors (and observers) to benzene and total gasoline vapors.
9. *Moen et al., 1995:* to evaluate the need for improvement of the working environment to reduce the risk of occupational cancer.
10. *Nordlinder & Ramnäs, 1987:* to investigate the exposure to benzene at different work places in Sweden where petrol is produced or handled.
11. *Runion, 1988:* to evaluate occupational exposures to potentially hazardous agents in the American petroleum industry.
12. *Nelson, 1995:* describes a measurement program to toxic organic compounds in Sydney's atmosphere.
13. *Wadge, 1996:* to assess the risks to health of non-occupational exposure to benzene in Australia.

Measurements reported in the company occupational hygiene reports were collected within the companies in the cohort (sources 1). In addition, exposure monitoring data to assess exposure to benzene for the urban and rural background values were from within Australia (sources 12 and 13). All the other sources for exposure data were collected outside the cohort (sources 2, 3, 4, 5, 6, 7, 8, 9, 10 and 11).

The sampling strategies that underlaid the measurement programme were:

1. No reports were supplied with measurements of participating petroleum companies, only spreadsheets with measurement results. However these data were supplied by the company hygienists who were supplied with *a priori* criteria, as to suitability of data. They also participated in a number of meetings over 2 years, agreed to supply all data that had been routinely collected and were typical of exposures and went through the data used to ensure that it was used appropriately.
2. The report summarizes results of exposure measurements to benzene in various types of jobs against the 8-hour time-weighted average 1 ppm "Action Level" and 5 ppm "Limit Value" over the period between 1977 and 1985.
3. The report summarizes measurements taken by CONCAWE member companies in Western Europe over the period 1986-1992 to assess exposure to benzene vapour during a range of refinery, distribution and retail activities. It updates CONCAWE report no. 3/86 (source 2).
4. The report provides information on measured human exposure data for occupationally exposed groups, consumers and the general public. Both 8-hour time-weighted average and short term exposure data are presented.
5. Personal breathing zone samples were collected during vessel cleaning operations in six small vessels and one large vessel to evaluate the adequacy of respirator selection.
6. Personal breathing zone benzene measurements were taken for tank truck drivers during loading, driving and unloading the gasoline. About half of the drivers were monitored during the loading operation at racks in a refinery and the other half was monitored for an entire trip between the refinery terminal and the gasoline station, including loading and unloading operation. Special references were made to the comparison of top and bottom loading.

7. Gasoline vapor concentrations were measured at a high volume service station in eastern Pennsylvania for on week in May 1983, for service station attendants, self-service customers and for various area locations.
8. Benzene and total hydrocarbon exposures were measured during removal, cleaning, pumping, and testing of underground gasoline storage tanks at five gasoline retail outlets in New Jersey during 1988.
9. Benzene and organic lead compounds were measured on eight Norwegian product tankers during loading and unloading in three Norwegian harbours.
10. Personal benzene exposure samples were taken at different workplaces in Sweden where petrol is produced or handled.
11. Occupational exposures to potentially hazardous agents were measured in the American petroleum industry during various jobs.
12. Thirty-six compounds classified as air toxics (including benzene) were collected in air samples at a variety of sites in the Sydney region in the period between 1976 and 1993.
13. This paper describes benzene exposure data in ambient air and in vehicles in Sydney, Melbourne, Brisbane, Adelaide, near to an oil refinery, and at a sample of large service stations in Australia.

Measurements (personal, static) that were taken:

1. Personal benzene measurements were taken by the occupational hygienists of the participating petroleum companies.
2. Personal task-based benzene measurements were taken from various jobs: service station attendants (N=546), road tanker drivers (N=238), road tanker drivers during top loading (N=1043), road tanker drivers during bottom loading (N=79), driver during delivery at service stations (N=22), rack operators and supervisors (N=89), rail car loaders (N=186), marine loaders (N=175), deck crew during closed marine loading (N=63), jetty staff during marine loading (N=98), refinery maintenance workers (N=134), refinery operators (N=692), refinery “off-site” workers (N=119), drum filler (N=15), octane rating laboratory workers (N=14), fuel testing laboratory workers (N=18).
3. Personal task-based benzene measurements were taken from various jobs: refinery operators (N=449), refinery off-site workers (N=426), refinery maintenance staff (N=55), tank drivers top loading (N=186), rack operators & supervisors (N=40), rail car loaders top loading (N=183), deck crew marine loading (N=19), jetty staff (N=92), service station attendants (N=82), service station cashiers (N=24), miscellaneous (N=193).
4. Personal benzene 8-hour time-weighted average exposures are measured for various jobs: on-site operator (N=847), off-site operator (N=819), maintenance worker (N=98), drum/barrel filler (N=9), road tanker driver (N=137), road tanker driver top loading (N=346), road tanker driver bottom loading (N=128), rack operators and supervisors (N=90), rail car operator top loading (N=215), jetty staff (N=113), deck crew (N=47), bridge crew (N=7), marine loading (N=11), terminal operator (N=53), service station attendant (N=1599), cashiers (N=24), airport operator (N=7). For some of these jobs short term measurements were also available.
5. Sixty-six personal breathing zone samples were collected by adsorptive air sampling in accordance with NIOSH Method 1501 for aromatic hydrocarbons. The duration of personal sampling was as long as possible (28-94 minutes), but was limited by the period of time that workers spent inside the vessels.
6. In total, 48 tank truck drivers were studied. Personal measurements were collected using diffusive samplers with carbon cloth KF-1500 as absorbent. Samples were analysed for benzene, toluene, xylenes and ethylbenzene.

7. In total, 45 personal air samples were collected (18 full shift and 27 short term measurements) on five consecutive days using 100/50 mg NIOSH-approved charcoal tubes.
8. Personal air samples were collected using 3M 3500 Organic Vapor Monitors with sampling times ranging from 15 minutes to 6.5 hours. Two laborers and an observer were monitored at each location.
9. Personal benzene and organic lead exposure was measured with activated charcoal tubes (SKC 226-01) by using low flow pumps (100 ml/min).
10. Personal benzene measurements were taken using sampling with Tenax tubes and personal sampling pumps for different categories of refinery workers: crude oil distillation unit (N=4), reformer unit (N=3), sampling and cleaning at laboratory (N=3), product analyses at laboratory (N=1), power plant workers (N=1), rail tanker loading top loading (N=21), rail tanker loading bottom loading (N=18), ship loading (N=11).
11. This paper gives an overview of personal benzene exposure measurements during various jobs and presents average exposure levels per job.
12. Thirty-seven stationary air samples were collected at 5 sites in Sydney at a flow rate of 500 ml/min for 5 minutes and then at 100 ml/min for 10 minutes. By choosing the sites, local sources such as roads were avoided. Samples were analysed for several compounds (including benzene).
13. Stationary ambient air samples were collected, which were analysed for benzene. In total, 860 measurements were collected in Sydney, 79 in Melbourne, 4032 in Brisbane, 47 in Adelaide and 92 near to an oil refinery, and at a sample of large service stations.

The background information collected regarding the measurements:

1. No reports were supplied with the exposure data, but the spreadsheets contained information on company, site, area, job title, date, duration of measurement, analytical method, product, % benzene, technology, proportion of time spent for that job. However, the team had free access to Shell Terminal exposure reports and were able to verify the exposure measurements, collecting for example the products handled and checking for exposure data, particularly from earlier periods, that had not been entered in the databases, which were based at the refineries. A small amount of data was identified.
2. Background information varied by measurement range described in the summary report. Information that was reported for every measurement series was: number of samples, country, the report or paper where data originated from. Some studies reported on meteorological conditions during the measurements.
3. Background information varied by measurement range described in the summary report. Information that was reported for every measurement series was: number of samples, year of sampling, and a process description.
4. Background information was reported about the job group, region, number of samples, and time period.
5. Information was collected about the type of vessel, sampling duration, performed task, the kind of respirator protection used.
6. Background information was obtained about the task performed and measurement duration. No information was reported about the product being loaded and unloaded.
7. Background information was reported on the job and sampling time. Meteorological test information was collected during the sampling period: wind speed, wind direction were monitored continuously and temperature, relative humidity and barometric pressure were acquired hourly from the Philadelphia International Airport and/or Willow Grove Air Station, both of which are within 30 miles of the service station.
8. Information was obtained about sampling time and job task.
9. Information was collected about the ship, the job title and task.
10. Background information was obtained about the performed task and number of samples.

11. A job description is given and information is presented about the performed task, the work area, and the agent that is determined.
12. Background information was obtained about the site where samples were collected.
13. Information was collected about the site and date of sampling, the measurement method and the sample size.

The method that was used for this study was an extension of that developed for the IOL study and used by the IP study. However, Health Watch had a high participation rate, good job descriptions from living study subjects and exposure condition information from co-workers. In addition, the work histories do not extend as far back in time as the IOL and IP work histories. These factors allowed a greater precision in allocating the tasks carried out to subjects in the study and therefore allowed a task-based approach to be used.

Jobs were normally made up of one or more activities, each of which had one or more tasks. The base estimate (BE) of exposure for a particular task was obtained from the arithmetic mean of occupational hygiene measurements taken by the petroleum companies or from data in the literature. If the BE related to a different exposure scenario than that being assessed, the BE required adjustment by one or more exposure modifiers (K-factors) to reflect the difference in: workplace (K_W), task (K_T), environmental (K_E) and materials (K_M) based on the IOL study.

The combination of tasks and K-factors were specific to each job title, site and time period.

The task estimate (TE) was then given by: $TE_{ijk} = BE \times K_1 \times ... \times K_s$. The task estimates were summed to give a time weighted average activity estimate (AE) for each job activity: $AE_{jk} = \sum_{i=1}^{n_{jk}} TE_{ijk} \times \left(T_{ijk} / A_{jk}\right)$. A similar time-weighted average calculation is done for the workplace estimate (WE) of exposure, normalised to a 35-hour week: $WE_k = \sum_{j=1}^{n_k} AE_{jk} \times \left(A_{jk} / 35\right) = \sum_{j=1}^{n_k} \sum_{i=1}^{n_{jk}} TE_{ijk} \times \left(T_{ijk} / 35\right)$. A new job start when the mix of activities or tasks changes or when a K-factor changes.

No (un)certainty score (as was used for the IOL study) was used to characterise the uncertainty in the work history lines or base estimates assigned to a work history line. In the executive summary ‘Retrospective Exposure Assessment for Benzene in the Australian Petroleum Industry’ (March 1998), some base estimates were described as being estimates with high uncertainty, because there were only few data available for that task.

A3.7.4 Assigning exposures to a Job-Exposure Matrix (JEM)

Exposure assessment was only carried out for inhalation exposure. It was decided not to develop a skin exposure rating for benzene following the IOL and IP studies.

No biomarker data were used.

Base estimates were assigned to each task within a job as reported in table 3 in the exposure assessment paper (Glass et al., 2000). A copy of this table is presented in Table A3.1 below.

Table A3.1
Base Estimates of Exposure (Glass et al., 2000)

Task	Base Estimate Benzene (ppm)
Airport background	0.08
Area 2 - Refinery A	0.14
Area 3A&B - Refinery A	0.23
Barge loading	2.21
CCU	0.16
CDU	0.11
DAP General work pre ‘89	1.86
DAP Head operator pre ‘89	0.74
DAP Maintenance	1.02
Dewatering	0.63
Driving&Unloading	0.16
Drum filling: stub, enclosed	3.90
Drum filling: stub/spear, open	3.52
Drum filling: stub, LEV	1.55
Drum laundry	0.39
Drum preparation	0.14
Gauging	4.20
Instrument fitter	0.48
Interceptor cleaning	0.12
Lab bench high	0.74
Lab bench low	0.15
Lab other high/low	0.09
Lab washing glassware	0.40
Mechanic	0.33
Mogas blending	0.42
Offshore operators	0.02
Onshore operators	0.06
Pigging	4.20
Rail car loading	3.77
Refinery fitter	0.38
Refinery operator, not exposed	0.07
Refinery operator plantwide	0.08
Reformer	0.39
Refuelling with Avgas	1.65
Road tanker bottom loading	0.55
Road tanker top loading	1.76
Rural background	0.001
Sampling	0.67
Separator skimming	0.12
Ship dip/gauge	5.41
Ship loading/unloading	0.11
Sour water	0.06
Tank cleaning 1	0.15
Tank cleaning 2	0.30
Tank cleaning 3	2.01
Tank farm – Refinery	0.15
Tank farm – Terminal	0.36
Terminal fitter	0.67
Terminal operator NE	0.14
Upstream fitter	0.04
Urban background	0.005

Subsequently adjustment factors (K-factors) were assigned to each base estimate to adjust for deviations from the situation in which base estimate measurements were collected: workplace (K_W), task (K_T), environmental (K_E) and materials (K_M) based on the IOL study.

Table A3.2
Schematic overview of available measurement data

Task	1950	1955	1960	1965	1970	1975	1980	1985	1990	1995
Airport background							Company data			
Area 2 - Refinery A							Company data			
Area 3A&B - Refinery A						Company data				
Barge loading							Company data			
CCU							Company data			
CDU							Company data			
DAP General work pre '89							Company data			
DAP Head operator pre '89									Comp. data	
DAP Maintenance							Company data			
Dewatering							Comp data	CONCAWE 1986	CONCAWE 1994	
Driving & Unloading							CONCAWE 1986	Kawai et al. 1991		
Drum filling: stub, enclosed							Company data			
Drum filling: stub/spear, open							Company data			
Drum filling: stub, LEV							Company data			
Drum laundry							Company data			
Drum preparation							Company data			
Gauging / Pigging								Company data		
Instrument fitter							Company data			
Interceptor cleaning							Company data			
Lab bench high							Company data			
Lab bench low							Company data			
Lab other high/low							Company data			
Lab washing glassware									Comp data	
Mechanic							Company data; CONCAWE 1986; Kearney&Dunham 1986; Nordlinder&Ramnas 1987			
Mogas blending										Comp data
Offshore operators									Comp data	
Onshore operators									Comp data	
Rail car loading							Company data			
Refinery fitter							Company data			
Refinery operator, not exposed							Company data			
Refinery operator plantwide							Company data			
Reformer						Company data				
Refuelling with Avgas								Company data		
Road tanker bottom loading							Company data			
Road tanker top loading							Company data			
Rural background									Nelson 1995; Wadge 1996	
Sampling								Company data		
Separator skimming								Company data		
Ship dip/gauge							Company data; CONCAWE 1986, 1994; Moen et al. 1995; Nordlinder&Ramnas 1987			
Ship loading/unloading								Company data		
Sour water								Comp data		
Tank cleaning 1										Comp data
Tank cleaning 2										Comp data
Tank cleaning 3										Comp data
Tank farm - Refinery							Company data			
Tank farm - Terminal							Company data			
Terminal fitter							Company data			
Terminal operator NE										
Upstream fitter									Comp data	
Urban background									Nelson 1995; Wadge 1996	

The schematic overview in Table A3.2 shows the available measurement data for the exposure assessment in this study.

The majority of measurements were collected in the period between 1980 and 1995. For some tasks, limited data were available and the research team described the base estimates based on those few data points as being uncertain (shaded in the diagram).

Detailed information was available from site visits and interviews with co-workers about the technology and work practices in use at various time periods and sites for subjects in the study. Therefore, modifying factors (K-factors) had to be applied to the base estimate to correct for those changes in technology and work practice in different time periods. Some of these factors were readily quantifiable, such as the benzene content of different products, but some were the result of professional judgement and were assessed by occupational hygienists from the Australian petroleum industry. These modifying factors are listed in Table A3.3.

Table A3.3
Value and frequency of use of technology modifiers for base estimates.

Task	Technology associated with task	Technology associated with base estimate	K-factor	Number of times K-factor applied
Dewatering	Closed drain	Open drain	0.2	12
Drum filling	Stub, displaced vapour	Stub, no LEV, open filling	1.0	17
Tank gauging	Remote gauging	Dip tape gauging	0.2	10
Tank gauging	Side gauging	Dip tape gauging	0.2	2
Jetty work	Chiksan arm	Flexible hose	0.5	13
Mechanical work	Gasoline solvent	Non-gasoline solvent	20.0	12
Rail car cleaning	Rail car cleaning	Gas test, scrape and hose	1.0	1
Rail car loading	Top splash fill and dip stick	Spear fill and dip stick	3.3	1
Sampling	Closed sampling	Open sampling	0.2	35
Tank cleaning	Gas test	Gas test, scrape and hose	0.14	3
Tank cleaning	No gas test, scrape and hose	Gas test, scrape and hose	1.0	4
Tank loading	Top splash	Spear, dip stick	3.0	1

All time periods for which no exposure measurements were available (see Table A3.2), had to be extrapolated based on information sources, which are listed in section A3.7.2. The exposure estimates were increased pre-1975 to take account of changes to sites as a response to environmental regulations on hydrocarbon emissions to reduce fugitive emission. Terminal fitters' exposure was also considered to have been exposed to more benzene before 1975 compared to after this date because of changes in work practices.

The other change that the exposure estimation methodology took account of was change in the task frequency eg number of dips made as a result of changes to customs requirements or changes to the range of tasks carried out by operators eg multiskilling meant that operators looked after more than one refinery unit. The job specific questionnaires asked about time on each task, technology and products.

A3.8 ESTIMATION OF INDIVIDUAL EXPOSURES

Cumulative exposure (in ppm-years) was calculated by multiplying the workplace exposure (WE) by length of time in a job and summed over jobs. In addition, exposure intensity, duration of employment, and exposure to benzene concentrate (yes/no) were used in the analysis. Estimates of unusual high exposure events eg skin and inhalation exposure from spillages, double fills etc, were also made and added to the exposure estimates for a separate analysis.

Ten out of the 390 subjects (=2.6%) had incomplete job histories. For these subjects, their date of first employment in the petroleum industry was known but no information was available on their early jobs. It was assumed that the subject had been doing their first known job since their date of first employment.

Where a subject had held more than one job, but information was not available about when the second job started, the period of employment was split equally between the two jobs.

In a few cases, the hours of work were missing so 40 hours per week was used as a default until 1973; from 1974 the basic working week was assumed to be 35 hours, in line with a union agreement.

A3.9 STATISTICAL ANALYSES

The final data set for the case control study is stored in a linked set of Microsoft Access database tables. Documentation for these is extensive and detailed, and we were supplied during the visit with copies of this documentation.

For the analysis phase, data were extracted by the Data Manager from the project database, either as text files, or as STATA native format files. Statistical analyses were carried out by the project statistician, using the statistical packages STATA. Odds ratios were calculated, comparing cases with their matched sets of controls. After exclusion of individuals with inadequate data, there were 79 cases, including 33 with leukaemia, and a total of 395 controls (165 for leukaemia cases).

The files of input instructions that drove the analyses in STATA, and the output logs that document when the jobs were run and whether they completed successfully, are all stored and can be inspected. Log files have also been kept for all analyses that were run interactively. Documentation still exists of the construction of graphs, and of the recalculation of different exposure metrics from the outputs of the exposure assessment. All log files are electronically date-stamped.

A number of matched analyses were performed, calculating odds ratios amongst subgroups defined in different ways:

1. start date in the industry (3 periods)
2. duration of employment (5 groups)
3. exposure to benzene in concentrated form (binary yes/no)
4. exposure intensity of highest exposed job (7 groups)
5. lifetime cumulative benzene exposure (6 geometric groups)

In each case, odds ratios were calculated in comparison with the lowest exposure category.

Separate analyses also fitted a conditional logistic model to cumulative exposure as a continuous variable.

No association was found between risk of leukaemia and either smoking or alcohol consumption, and no adjustment was made for these in the analyses by exposure. There was also no relationship with date of starting work or duration of employment.

No association was found between any exposure measure and risk of Non-Hodgkin lymphoma or multiple myeloma. However, leukaemia cases had, on average, a higher lifetime exposure than controls, and a greater proportion of cases were in higher exposure categories. Analysis by binary exposure to benzene concentrate showed a significant odds ratio of 12.5. Analysis by geometric groupings of cumulative exposure suggested an excess in the highest exposed group, and some evidence of an increase in the middle groups over the lowest; analyses by average intensity of exposure showed very similar patterns. Some analyses fitted more than one of these exposure metrics, but the authors pointed to the difficulty of interpreting these because of correlation between the metrics.

The analysis by continuous exposure yielded an odds ratio of 1.65, or an increase in risk of 65%, per doubling of exposure (95% CI 1.27-2.17). Cross-validation of the geometrically grouped model was carried out by grouping on quintiles of exposure, yielding similar results.

Additional analyses were carried out on subgroups of the leukaemias defined by cell type. These were necessarily based on small numbers, but suggested that there were increased risks associated with cumulative exposure, for ANLL (11 cases) and CLL (also 11 cases), but not for CML (9 cases).

Power calculations were not carried out for the main analyses, principally because the results appeared to show a positive association of risk with exposure. It was clear that power would have been low in analyses of the individual cell types, but again calculations were not carried out.

A3.10 VALIDATION AND RELIABILITY

Validation data with which to compare the base estimates (BEs) were sought in the occupational hygiene literature including reports from CONCAWE and non-Australian petroleum industry sources and did not include sources already used for constructing BEs. The BE for a particular task was considered to have been confirmed for the purposes of the study, if it was within 20% of the mean of the validation data sets for the task, or within the interquartile range (IQR). Where neither of these conditions was met, the BE was adjusted.

Four out of fifty-one BEs (=8%) were changed as a result of this exercise, two of these were literature-based BEs and two changes resulted from the exclusion of outliers.

A3.11 CHECKING RANDOM SAMPLE FOR ERRORS OR INCONSISTENCIES

A random sample of cases and controls were checked for errors or inconsistencies. Three cases sets were randomly selected from the case-control data set (Case Set 1, 55, and 78). Each case set consisted of one case and their accompanying controls. For this data checking the case and three of their accompanying controls were selected and checked with the Microsoft Access database on: company, site, area, job title, and time period. All information in the database matched the written work histories, except for one work history line for one of the controls. The line was split into ‘operator’ and ‘engine driver’, for which no motivation could be found in the written work history. Because both ‘operator’ and ‘engine driver’ were given the same base estimate, this did not influence the results in this study.

APPENDIX A4: REVIEW OF REPORTS OF API STUDY

A4.1 INTRODUCTION

Between 1984 and 1985 a feasibility study was carried out by Environmental Health Associates, inc. of Oakland, CA, to determine whether it was feasible to carry out a retrospective mortality study of employees of the petroleum industry exposed to downstream gasoline (petroleum). It was judged feasible, and the study began at the end of 1985, carried out by ENSR Health Sciences of Alameda, CA.

The PI of the feasibility study, the subsequent cohort study and the eventual nested case-control study was Otto Wong. When this review project began, CONCAWE contacted Dr Wong, but he had left the employment of the research bodies, and had no residual contact with the research materials. Since it was not clear where the materials now resided, nor who was their current custodian, we agreed with CONCAWE that it was not possible to make a site visit, or to attempt audit of any of the research materials or documents. This review is therefore based entirely on the published papers and final reports available to the review team.

A4.2 COHORT DEFINITION

The API case-control study was nested in a cohort study of petroleum (gasoline) workers in the USA. Following a feasibility study, the original cohort study was commissioned in 1984. There were two separate cohorts: marketing and distribution employees at land-based terminals, and marine distribution employees who were exposed to gasoline. Two companies contributed data to both cohorts, a third to the land-based cohort only, and a fourth to the marine cohort. The case-control study was nested within the land-based terminal cohort only.

The data collection team, consisting of a supervisor and several research assistants, visited more than 30 record locations where employment records were kept. Personnel files were screened to ensure the selected workers had opportunity for gasoline exposure (interpreted liberally). The authors note that since this judgment was made by local contacts, even with some measures to improve standardisation, it is possible that some jobs with low exposures might have been omitted at some sites.

On the basis of the exposed jobs at each centre, employment records were extracted and microfilmed. In cases of doubt, records were copied, leading to “a large number of records with no or unknown exposures). Over two years, around 68,000 employment records were microfilmed.

From these records, data were entered including identification data, race, sex, dates of birth and hire, termination or retirement date, employment status, vital status and date of death (if known). A summary work history consisted of jobs in chronological order, with dates, jobs and locations. For some incomplete records, additional information was sought from the companies.

The feasibility study had concluded that a complete verification of the cohort was not possible, but several attempts were made to check for possible sources of systematic bias, including cross-checking death data from different sources. It was concluded that there was no evidence for wholesale gaps in the cohort data set, and that if there were missing records, they must be small in number.

An exposure assessment exercise was carried out base on measurement data from 600 industrial hygiene samples, and classifying jobs into four categories:

- Drivers

- Loaders
- Terminal operators
- Other terminal jobs.

In the land-based cohort, 4,800 job titles were collapsed into these four categories. The job-exposure matrix was stratified into four time periods for truck operations. Indices were calculated of individuals' cumulative exposure to Total Hydrocarbons (THC), mean intensity of exposures, duration exposed and peak exposures.

The assembled cohort for the land-based terminals numbered 9,026, mostly white males, who met the entry criterion of at least one year's potential exposure. More than 80% were terminated or retired before the study began. Vital status was determined as of June 30, 1989. About a quarter were found to have died, and death certificates were obtained for over 97% of these.

SMRs were calculated using US national rates, for all causes and for a range of non-cancer and cancer causes. For missing death certificates, causes of death were redistributed according to the proportions in the observed certificates. The results showed the usual "healthy worker effect", and the SMRs showed no relationship with any of the exposure indices, grouped into exposure categories.

Cox proportional hazard modelling was carried out for kidney cancer and for leukaemia, treating cumulative exposure, peak exposure, duration and intensity as continuous predictors. (From the descriptions given, it seems that these were not calculated as time-dependent variables.) Again, no significant relationship was observed with any of the exposure indices.

A4.3 CASE DEFINITION

The original plan for the cohort included additional case-control investigations for causes of interest. The diseases initially chosen included total leukaemia, acute myeloid leukaemia, multiple myeloma (because of benzene in gasoline), and kidney cancer (because of animal studies).

Selection of cases was based on the causes of death from the death certificates obtained for the land-based cohort, either as underlying or as a contributory cause. This yielded 35 cases of leukaemia, including 13 of acute myeloid leukaemia; 11 cases of multiple myeloma; and 12 cases of kidney cancer.

All data were collected from the cohort records, and inclusion in the case-control study required no active participation from the cases.

A4.4 CONTROL DEFINITION

Controls were randomly selected from the cohort study database, to match individual cases at a target ratio of 5:1. For each case, the controls were to be selected from the same company, and the same sex, and to have a date of birth within two years of the case; and they must have been alive at the time of death of the corresponding case.

All data were collected from the cohort records, and inclusion in the case-control study required no active participation from the controls.

A4.5 DATA COLLECTION IN CASES AND CONTROLS

The data held for cases and controls in the cohort database included

- Name
- Social Security Number
- Study identification number
- Race
- Sex
- Birth date
- Hire date
- Termination or retirement date (if applicable)
- Summary work history
- Employment status
- Vital status and date of death.

There were no data on smoking habits or other risk factors of lifestyle or previous employment.

After cases and controls were selected, no new data were collected from the cases and controls for this nested case-control study.

A4.6 WORK HISTORIES IN CASES AND CONTROLS

Work history data from the cohort database included, for each job held:

- Date of job change
- Job location
- Job title

These data had come direct from the employment records, and there was no use of proxy respondents in their collection.

The reports on the cohort study record that, in a few cases where employment records were incomplete or uninterpretable, additional clarification was sought from the companies. They do not report how successful this was, nor what is meant by "a few". There is no description of whether any data loss might be systematic or period-specific. However, they give the impression that the effect of missing data was likely to be negligible.

A4.7 EXPOSURE ASSESSMENT AND ASSIGNMENT

A4.7.1 Background

Exposure measurements from 1975-1985 were used to quantify exposure intensities associated with specific combinations of job tasks, types of worksites, and types of products. Based on the source-receptor concept, the type of work site and the configuration of equipment at the site were considered important determinants of job exposures. Each work site had a unique history of changes in operations and equipment, which may have affected exposures. Therefore, facility histories were sought for the work sites identified in job histories from each of the companies.

Extensive information on recent and past truck and marine operations was collected by site visits, interviews of long-term employees and annuitants from each of the participating companies, and company completion of facility questionnaires on the history of equipment and operations at specific terminal sites and on selected marine vessels. These data were combined with published reports of industry-wide activities to develop job descriptions of tasks, work-site descriptions of typical operations, and historical changes across the industry.

From this information, four factors were obtained for facilities at truck terminals: splash or submerged top loading, metered or valved top-loading controls, bottom loading, and presence of a vapour recovery system. For marine vessels, four types of data were obtained: loading with hatches open or with remote venting, voyage frequency, percentage of gasoline in cargo, and area of operations.

These data on individual terminals and vessels allowed the individualization of exposure estimates for subjects who worked at these sites.

A4.7.2 Exposure measurements

Exposure data collected by the four participating companies between 1975 and 1985 were used as the exposure assessment base. A small number of measurements were collected by the University of Massachusetts to fill data gaps and verify earlier observations.

Table A4.1
Short-term gasoline vapour exposures during loading and delivery tasks (1975-1985)

Facility type	**N**	**THC arithmetic mean (ppm)**	**SE**	**Range of sample duration (min)**
Truck loading				
No vapor recovery	139	130	14.2	7-60
Top loading (3 companies, 8 sites)	103	120	15.4	7-60
Bottom loading (1 company, 1 site)	36	157	32.4	8-60
Vapor recovery	81	17	2.8	8-60
Top loading (1 company, 1 site)	42	11	1.1	8-60
Bottom loading (2 companies, 6 sites)	39	24	5.6	15-40
Truck deliveries				
Large deliveries: remote vented, large underground tanks using tight connections (1 company, 7 sites)	32	9	1.8	15-30
Small deliveries: drivers from small terminals, some above-ground tanks and confined locations (OPA, 1988)	24	400	NR	15

NR = not reported
OPA = Ontario Petroleum Association

Approximately 600 samples had been collected by several methods and analysed for total hydrocarbon (THC) concentration. Data on benzene exposure levels were not available. Both short-term (15-90 minutes; Table A4.1) and full shift personal samples (Table A4.2) were collected as well as fixed location area samples to measure low-level background exposures.

Table A4.2
Time weighted average, full shift exposures for truck operation jobs by type of facilities (1975-1985)

Generic job / terminal loading facility	N	THC arithmetic mean (ppm)	SE
Driver			
No vapor recovery	98	14	1.5
Top loading (3 companies, 12 sites)	90	14	1.5
Bottom loading (2 companies, 2 sites)	8	18	2.4
Vapor recovery	94	9	1.6
Top loading (1 company, 1 site)	7	10	1.5
Bottom loading (3 companies, 17 sites)	87	9	1.6
Terminal operator			
Top load, no vapor recovery (3 companies, 12 sites)	37	9	1.3
Vapor recovery (3 companies, 41 sites)	112	5	0.8
Other terminal jobs	14	5	1.3

None of the exposure measurement results, which were used for the exposure assessment, were referenced in the exposure assessment report or paper (Smith et al., 1993) and were not available for this review. Therefore, we do not know what the measurement strategy was and what background information was collected regarding those measurements.

Given an estimate of average exposure intensity for each task with exposure by work-site type and time spent on the tasks, a time weighted average (TWA) exposure could be estimated for each job title and work-site type by using a task-TWA exposure model:

$$task-TWA = \frac{\sum_{j=1}^{all\ tasks}\left[(task\ mean)_j (task\ time)_j\right]}{\sum_{j=1}^{all\ tasks}(task\ time)_j}$$

Past exposures of gasoline distribution workers were extrapolated with the task-TWA exposure model based on major changes in job definitions and work-site characteristics that had occurred across the industry, for which information was derived from the sources described in section A4.7.1. Changes in worker behaviour over time were also used to modify exposure, but it was not explained how this was done and where this information was derived from.

Four time periods with distinct characteristics were identified for truck operations: pre-1950, 1950-1964, 1965-1974, and 1975-1985. Although sharp transition dates are given, they represent median dates of system changes. A summary of extrapolated, full shift, time-weighted average THC exposures for generic jobs and specific site types by time period is given in Table A4.3. How these exposure estimates were extrapolated is not described in the available information.

Table A4.3
Summary of extrapolated, full shift, time weighted average exposures to total hydrocarbon (ppm) from gasoline for generic jobs and specific site types by time period.

	Terminal size						
	Large				*Small*		*Both*
Year and load type	**Driver**	**Driver/loader**	**Loader**	**Terminal operator**	**Driver**	**Terminal operator**	**Other**
1975-1985							
Submerged	14	7	62	9	-	-	8
Splash	-	-	-	-	180	72	-
Vapor recovery	9	6	8	5	-	-	3
1965-1974							
Submerged	64	41	63	29	-	-	13
Splash	79	42	97	34	180	72	17
Vapor recovery	41	39	10	25	-	-	8
1950-1964							
Submerged	190	150	98	72	-	-	22
Splash	220	150	150	80	210	80	28
Pre-1950							
Splash	170	140	75	68	170	68	19

A4.7.3 Assigning exposures to a Job-Exposure Matrix (JEM)

Exposure assessment was carried out for inhalation exposure only. No estimate for dermal exposure potential was assessed and no biomarker data were used.

Average daily exposures were assigned to each subject based on job title and work site characteristics for each work history line. Hundreds of different job titles appeared on the job histories of cohort subjects, submitted by the participating companies. It was not feasible to estimate a specific exposure for each job title and many of the titles were synonyms, abbreviations, or misspelled versions of common titles. Therefore, each company assembled a group of knowledgeable individuals (such as annuitants, active personnel, and industrial hygienists) to review the job titles and form a small set of generic jobs.

For truck operations, four generic job groups were identified based on their potential for work around emission sources of gasoline vapours: driver (tasks: loading trucks, driving, making deliveries, and other non-exposed tasks), loader (tasks: truck loading or non-exposed tasks), terminal operator (tasks: some loading, a variety of mechanical and maintenance tasks, and non-exposed tasks), and other terminal jobs. Other terminal jobs included all jobs with no potential for direct contact with emission sources, and were therefore exposed only to background levels of vapour in the terminal area.

For marine operations, two generic job groups were identified: deck personnel and other shipboard jobs. However, because only few workers were classified in one of the marine job groups, the case-control study was limited to the land based workers only.

From the limited information that was available for this review (report and published paper (Smith et al., 1993)), there is no indication of any gaps in knowledge about job histories or exposure measurements to fill the job exposure matrix, so we don’t know whether imputation was required.

A4.8 ESTIMATION OF INDIVIDUAL EXPOSURES

Cumulative exposure to total hydrocarbon (in ppm-years) was calculated by multiplying the TWA exposure for each job in a subject’s job history by the duration in the job and summing across all jobs.

In addition, the annual frequency of peak exposures was determined by identifying the tasks with potential to produce peaks. A peak exposure was defined as at least 500 ppm THC averaged over 15-90 minutes. The number of occurrences of a task was determined from the historical data.

A4.9 STATISTICAL ANALYSES

In all case sets, the exposure indices for controls were truncated at the time of death of the corresponding case.

The data were analysed, comparing cases with their matched controls, by the Mantel-Haenszel procedure, for comparisons between job categories (ever/never employed in) and year of first exposure. For continuous predictors such as years of exposure and cumulative THC exposure, univariate and multivariate conditional logistic regression procedures were used.

The authors do not report which statistical package was used to carry out the analyses.

For all leukaemia, and for the subtype AML (as well as for multiple myeloma and kidney cancer), univariate conditional logistic regression analyses were run using as predictors each of:

- Duration of employment
- Duration of exposure
- Cumulative exposure
- Frequency of peak exposure
- Year of first exposure (binary, split at 01/01/1949)

Models were also fitted with the following pairs of variables:

- Duration of exposure
- Cumulative exposure

- Duration of exposure
- Frequency of peak exposure
- Cumulative exposure
- Frequency of peak exposure

None of the models appears to have been adjusted for any other risk factor. Risks of leukaemia or of AML did not relate significantly to any job category, or to any of the exposure indices in either univariate or bivariate regression models.

No power calculations are reported.

A4.10 VALIDATION AND RELIABILITY

The authors note explicitly that there was no external validation possible for cohort completeness, because there was only one principal source of data, i.e. the employment records. For the same reason, the opportunities for verification of individual data items would be limited.

The authors describe a number of side-checks on parts of the cohort, and conclude from these that there is no reason to suspect large systematic gaps in the cohort ascertainment.

The accuracy of the task-TWA model was checked by comparing the 1975-1985 task-TWA estimate for drivers at the two most common terminal configurations (top loading vapor recovery and bottom loading with vapor recovery), relative to the full-shift exposures. The extrapolation appeared to be within +64% to -27% of the observed mean THC concentrations.

The bivariate regression models acted as sensitivity checks to the extent that they demonstrated very similar results to those from the univariate models, so that the null conclusions were not a result of negative confounding.

Institute of Occupational Medicine

World Health Organisation
Collaborating Centre
for Occupational Health

Applying science for a better working environment

The Institute of Occupational Medicine

The IOM is a major independent centre of scientific excellence in the fields of occupational and environmental health, hygiene and safety. We aim to provide quality research, consultancy and training to help to ensure that people's health is not damaged by conditions at work or in the environment. Our principal research disciplines are exposure assessment, epidemiology, toxicology, ergonomics and behavioural and social sciences, with a strong focus on multi-disciplinary approaches to problem solving.

Our beginnings

Our first major research programme began in the 1950s, on respiratory health problems in the coal mining industry. Major themes were quantification of airborne dust concentrations in different jobs, characterisation of types and constituents of the dusts, measurement of health effects, relationships between exposure and disease, and proposals for prevention. This research became an international benchmark for epidemiological studies of occupational health, and was the primary influence on dust standards in mines in the UK, US and other countries.

Current themes

Our current work spans many other industries including asbestos, MMMF, pesticides, chemicals, energy, telecoms, metals, textiles, construction, agriculture as well as the environment. While diseases of the respiratory tract remain a major interest, our scope now extends to many other health outcomes such as mortality, cardiovascular effects, cancer, back pain, upper-limb disorders, hearing loss, skin diseases, thermal stress and psychological stress. Related work includes the development and application of measurement and control systems, mathematical models and survey methods.

Who we work for

Our work in these areas is conducted for a wide range of organisations in the UK, the EU, and the US, including Government departments, international agencies, industry associations, local authorities, charitable organisations, and industrial and commercial companies. The IOM is a World Heath Organisation (WHO) collaborating centre and is an approved institute of the Universities of Edinburgh and Aberdeen, enjoying collaborative research links with NIOSH, IARC, and many other institutes throughout the world.

Publication

We believe that our research findings should be publicly available and subject to the scrutiny of the international scientific community. We publish our findings in the peer reviewed scientific literature and through our own series of Research Reports.

Contact

For further information about the IOM's research capabilities:

Dr Robert Aitken
Director of Research Development
Rob.aitken@iomhq.org.uk

Research Park North, Riccarton, Edinburgh, EH14 4AP, Scotland, UK
Tel +44 (0)870 850 5131 • Fax +44 (0)870 850 5132 • e-mail iom@iomhq.org.uk

A recognised charity limited by guarantee registered in Scotland No. 123972
Multi-disciplinary specialists in Occupational & Environmental Health and Hygiene • Research • Consultancy • Analysis • Training