METHODOLOGY

Reporting of occupational and environmental research: use and misuse of statistical and epidemiological methods

Lesley Rushton

Abstract

Objectives—To report some of the most serious omissions and errors which may occur in papers submitted to *Occupational and Environmental Medicine*, and to give guidelines on the essential components that should be included in papers reporting results from studies of occupational and environmental health.

Methods—Since 1994 *Occupational and Environmental Medicine* has used a panel of medical statisticians to review submitted papers which have a substantial statistical content. Although some studies may have genuine errors in their design, execution, and analysis, many of the problems identified during the reviewing process are due to inadequate and incomplete reporting of essential aspects of a study. This paper outlines some of the most important errors and omissions that may occur. Observational studies are often the preferred choice of design in occupational and environmental medicine. Some of the issues relating to design, execution, and analysis which should be considered when reporting three of the most common observational study designs, cross sectional, case-control, and cohort are described. An illustration of good reporting practice is given for each. Various mathematical modelling techniques are often used in the analysis of these studies, the reporting of which causes a major problem to some authors. Suggestions for the presentation of results from modelling are made.

Conclusions—There is increasing interest in the development and application of formal "good epidemiology practices". These not only consider issues of data quality, study design, and study conduct, but through their structured approach to the documentation of the study procedures, provide the potential for more rigorous reporting of the results in the scientific literature.

(*Occup Environ Med* 2000;**57**:1–9)

MRC Institute for Environment and Health, University of Leicester, UK
L Rushton

Correspondence to:
Dr Lesley Rushton, MRC Institute for Environment and Health, University of Leicester, 94 Regent Road, Leicester LE1 7DD, UK

As with all areas of research, research workers in the field of occupational and environmental health seek to publish their findings in respected journals such as *Occupational and Environmental Medicine*, both to disseminate their results and to give their work scientific credibility. In this area of medical research, as with others, statistical and epidemiological methods play an important part. The use of statistics in clinical research has been widely discussed, and it has been pointed out that their misuse is both unethical and can have serious clinical consequences.[1]

To encourage doctors and health professionals to use and understand statistical techniques, statistical guidelines, checklists, and articles on specific methods have been published.[2][3] The statistical content of published papers in different medical journals has also been the subject of extensive review. It has been shown that as many as 50% of papers are purely descriptive, with no statistical methods, and many restrict themselves to simple techniques such as the *t* test, contingency tables, and Pearson's correlation, although the percentage of papers with only these common techniques has fallen over the years.[4][5]

There have been many surveys of statistical errors in the medical literature with error rates ranging from 30% to 90%.[6–9] McGuigan[10] reviewed all 12 issues of the 1993 *British Journal of Psychiatry* for statistical errors and compared these with a previous review of the same journal in 1977–8.[11] There was little sign of an improvement between the two reviews. Error rates in the 1993 review varied from 1% in the description of the statistics used, through 27% of papers failing to present adequate summary statistics, to 80% for misapplication of the *t* test or lack of reporting the type of *t* test.

The aim of this paper is to draw attention to some of the most important omissions and errors that can occur and to give guidelines on the essential elements that should be included in papers reporting results from studies investigating occupational and environmental health.

Statistical and epidemiological reviewing in *Occupational and Environmental Medicine*

Since October 1994 *Occupational and Environmental Medicine* has used a panel of about 20

Table 1   Checklist for epidemiological and statistical reviews

Design:
  Is there a clear statement of the study objectives?
  Is there a clear description of the study design?
  Is the study design appropriate?
  In there a clear description of the study populations?
  Is there a clear description of how the main variables were measured?
  Is there an assessment of the adequacy of the sample size?
Presentation and analysis:
  Are the statistical procedures adequately described or referenced?
  Are the statistical procedures appropriate?
  Are the response rates reported?
  Are the data and the results adequately described?
  Are confidence intervals and significant levels given where appropriate?
Bias and confounding:
  Are the response or tracing rates satisfactory?
  Are you reasonably satisfied that the results are not explained by bias in subject selection or measurement?
  As far as you can tell, have confounding variables been taken into account appropriately—either in the analysis or in the discussion?

medical statisticians, one of which, as well as the normal two reviewers, will review papers which have a substantial statistical content.[12] As well as providing written comments for the authors of the paper the statistician completes a checklist, as shown in table 1. Reviewers tick one of: not applicable; yes; no; or unclear for each question on the checklist, and follow this by a recommendation that the paper is either acceptable, acceptable after minor revision with no need for further review, acceptable with revision and will need reviewing again, or not acceptable. This procedure is similar to those used for other journals and the checklists suggested by Altman.[13]

Just over 300 papers are submitted to *Occupational and Environmental Medicine* each year and about half of these are accepted for publication. About a third of submitted papers are reviewed by a statistician, usually after being reviewed by two other reviewers.

Important statistical errors that may occur

A wide variety of statistical and epidemiological errors and omissions can occur and some of the most serious are given in table 2. These are categorised into five main areas, similar to those used by Altman,[13] although many of his examples refer to intervention studies rather than observational studies, which are more

Table 2   Serious errors and omissions occurring in papers submitted to OEM

Design:
  Authors unclear of type of epidemiological study
  Adequacy of sample size not considered
  Bias in selection of subjects
Execution:
  Data collection problems and missing data not adequately reported
  Non-respondents not investigated
  Sample selection and exclusions inadequately justified
Analysis:
  Parametric tests carried out on obviously skewed data
  Use of multiple paired tests
  Inappropriate analysis of repeated measures or longitudinal data
  Incorrect analysis of matched case-control studies
  Modelling incorrect—for example, inadequate adjustment for confounders, interaction terms not included, only significant variables from preliminary analyses included
Presentation:
  Inadequate description of the methodology and statistical procedures
  Inappropriate summary statistics for non-normal data
  No presentation of risk estimates—for example, odds ratios—and confidence intervals
Interpretation:
  Potential bias due to sample selection, no or poor response, missing values, exclusions
  Lack of statistical power not considered
  No allowance made for multiple testing
  Misunderstanding and misinterpretation of results from models

commonly carried out in occupational and environmental research.

DESIGN AND EXECUTION
Although seemingly surprising, authors may be unclear about which epidemiological design they have actually used. A major concern in the design of studies is the almost universal lack of reporting on the adequacy of the sample size, and if it was chosen on the basis of power calculations. Many studies have problems with the selection of their subjects and inadequacies in this part of the execution of the study may be glossed over in the paper. Knowledge of exactly how the sample was selected, missing data, and exclusions is essential for assessing the generalisability of the results. Another area of concern is a lack of consideration of the healthy worker effect—that is, that many employed groups are likely to be healthier than the comparison population, which is often the national population. It is also worth remembering that, as Nemery *et al*[14] point out, the factories, workshops, or sites included in a study may not always be representative of the industry as a whole. Willingness to participate may indicate that conditions are reasonable, leading to a healthy workshop effect.

ANALYSIS
Table 2 presents some of the more severe errors which may be found in the analytical techniques, although there are many more minor errors which can occur. Basic errors include (*a*) failure to check assumptions inherent in the method of analysis, for example, normality of data when using parametric significance tests, (*b*) ignoring pairing or ordered categories and thus using an inappropriate statistical test, (*c*) seemingly arbitrary categorisation of continuous variables or categorisation to produce significant comparisons, and (*d*) repeated use of the χ^2 test for subdivisions of a large table, rather than as a test for an overall association.

More major errors include (*a*) multiple comparisons—for example, carrying out many *t* tests—thereby increasing the likelihood of a significant result, (*b*) ignoring the repeated design characteristics of data during the analysis stage—for example, if the data relate to continuous monitoring of a health end point, and (*c*) carrying out a non-matched analysis for a matched case-control study. Where authors really get into difficulty is when mathematical modelling techniques are used. It may sometimes be obvious that authors have an inadequate understanding of these techniques and would have been well advised to seek expert help. The ready accessibility of user friendly statistical computer packages from which output can be produced with ease may add to the temptation to try and use modelling without giving careful consideration as to whether it is appropriate and how to interpret it.

PRESENTATION
There seems to be a general lack of transparency in some papers in that, until pointed out by the reviewer, essential descriptions of the method, including the statistical methods, are

often omitted. Many textbooks give good advice on the presentation of results.[13 15] Although not essential, it is useful if the computer package is named, as many readers will be familiar with the type of output and also the potential problems with the packages (which all have their own quirks).

Despite articles and books[16] concerning the presentation of confidence intervals around point estimates, these may still be omitted when a paper is first submitted to a journal. Altman[11] also draws attention to (*a*) unnecessary (or spurious) precision in quoting results, often unfortunately, transcribed directly from the statistical package for which the default may be six or more decimal places, and (*b*) misleading features of graphical presentation, both of which occur often in papers submitted to *Occupational and Environmental Medicine*.

INTERPRETATION

Errors encountered in the design and analysis of a study can also continue through to errors in interpretation. Potential problems due to lack of statistical power, poor response, and bias in subject selection or data collection may not be considered adequately by the authors. Again the use of modelling can cause immense problems of interpretation. For example, there may be no presentation or assessment of the goodness of fit of model, and the use of interaction terms is either not considered or misinterpreted.

Misinterpretation of the results of significance tests can include (*a*) the suggestion that the smaller the p value is the stronger must be the effect, (*b*) any significant result is worthy of comment, even if the effect is small or unlikely, and (*c*) non-significance implies the proving of the null hypothesis.

Guidelines for reporting studies

This paper has referred to check lists and guidance already published and authors would find these extremely useful for general advice on the use and presentation of medical statistics. However, in occupational and environmental medicine, observational studies are often the preferred choice of design, on which previous guidance has not focused specifically. Also, many studies incorporate assessment of exposures as an essential element of data collection, an aspect which may have particular problems. It is not the aim of this paper to discuss how these studies should be carried out and authors should refer to the many good textbooks on general epidemiological methods.[17–21] This paper gives some guidelines and illustrations of the essential components of these studies which should be reported in papers, so that the reader may fully understand and interpret the research. The more complex analytical techniques of these studies use various modelling techniques so additional guidance is suggested on the reporting of these.

GENERAL ISSUES

As Campbell and Machin point out[15] classification of a research paper into the study type before detailed reading is useful to alert the reader to issues which may be unfamiliar to them. However, many areas are common to all study designs. Firstly, the objectives of all studies need to be clearly stated, including the variables to be measured, and the adequacy of the sample size should be justified.

The limiting factor in many epidemiological studies is often the errors encountered in measuring or assessing the exposures. One effect of this error is the potential to cause bias in the measures of association between the exposure and the health outcome. The possible misclassification of study subjects by disease or exposure, information bias, can result in misleading conclusions. A distinction needs to be made between (*a*) non-differential misclassification—that is, when the likelihood of misclassification is the same for both groups, and (*b*) differential misclassification—that is, when the likelihood of misclassification is different between groups—for example, between diseased and non-diseased people. The first tends to bias the effect estimate towards the null value.[22] For example, consider a hypothetical example from a study of self reported dermatitis in printing workers. Suppose the true prevalence rates are 30% in those exposed and 15% in those not exposed (table 3). The prevalence ratio is thus 2.00 and the prevalence odds ratio (OR) is 2.4. Suppose that 20% of those with dermatitis incorrectly report that they do not have dermatitis and 10% of those without dermatitis incorrectly report that they do have dermatitis. The prevalence ratio and prevalence OR are reduced to 1.5 and 1.7 respectively. Non-differential misclassification may thus lead to a lack of association between exposure or disease.

Differential misclassification can bias the observed effect estimate either towards or away from the null value. For example, in a case-control study of lung cancer, recall of a relevant occupational exposure may be different in the cases than in the healthy controls.

Most occupational studies of risk also involve some evaluation of confounding effects. A confounding factor is a variable which is (*a*) a risk factor for the disease of interest, even in the absence of exposure (either causal or in association with other causal factors), and (*b*) is associated with the exposure but is not a consequence of exposure.

Confounding can be controlled for in the study design—for example, by matching in a case-control study—or in the analysis by stratification or multivariate analysis. Deciding whether to adjust for a given variable is not always straightforward. Schlesselman[20] discussed the problems of carrying out signifi-

Table 3   Hypothetical data from a prevalence study of self reported dermatitis in printing workers

	True prevalence		Reported prevalence	
	Exposed	*Non-exposed*	*Exposed*	*Non-exposed*
Dermatitis	30	15	24+7 = 31	12+8.5 = 20.5
Non-dermatitis	70	85	63+6 = 69	76.5+3 = 79.5
Total	100	100		
Prevalence ratio	2.0		1.5	
Prevalence odds ratio	2.4		1.7	

cance tests for differences between cases and controls for all potential confounders. For example, he pointed out that even if a variable shows a significant case-control difference, adjustment may be unnecessary because of a lack of association between the variable and exposure. Alternatively a non-significant difference may none the less accompany a large change between the adjusted and unadjusted estimates of the effect of exposure. In general it is informative to present both unadjusted and adjusted risk estimates so that the reader can assess the effect of confounders.

Misclassification of a confounder tends to lead to under adjustment for the confounder—that is, overestimate the association between the exposure and disease outcome. The resulting estimates may also differ between strata creating a spurious interaction effect.

The following sections outline some of the essential points that should be explained when describing results from three types of study design. An illustration of good reporting practice is given for each, with the results from some of the mathematical modelling from the three examples being given in the later section discussing the reporting of mathematical models.

Cross sectional studies
In a cross sectional study all the information is collected at one time. The risk measure is that of disease prevalence rather than incidence, and cross sectional studies are particularly useful for investigating non-fatal degenerative diseases, often with no clear point of onset—for example, musculoskeletal problems. In reporting these studies the reader needs to know (*a*) the sources of and methods used to obtain information on health outcome—for example, physical examinations, self reported questionnaires, and medical records, (*b*) ascertainment of measurement of exposure—for example, industrial hygiene measurements, work histories, (*c*) how subjects were selected for inclusion, and (*d*) how a comparison population was chosen. All these may involve problems of inaccurate, incomplete, or conflicting information. Restrictions on inclusion criteria may have implications for the generalisability of the study.

Comparison groups can be internal, particularly when examining prevalence gradients according to exposure levels, or external. The problem of the healthy worker effect mentioned earlier can be even stronger in cross sectional studies than in cohort studies because they usually only include actively employed workers. In cross sectional studies both exposure and health outcome are determined simultaneously. Although an association can be shown between

exposure and health outcome, cross sectional studies are thus limited in their ability to establish causality, by contrast with other epidemiological designs in which the potential cause clearly precedes the health outcome.

Thun *et al*[23] reported results from a small cross sectional study which investigated the relation between cadmium exposure and kidney dysfunction in workers at a cadmium production plant. Nineteen actively employed production workers and 27 highly exposed, former workers, were compared with an unexposed group of 32, enrolled from a local hospital. The authors describe the questionnaire which elicited information about age and medical history, the measurement of physical characteristics—such as height, weight and blood pressure—and the selection of biomarkers to assess renal function. The authors discuss the potential for selection bias and the comparability of the hospital workers control group with the cadmium workers for education and socioeconomic status. They suggest that the non-exposed group provided referent values for the physical measurements and allowed better control for age, an important potential confounder for the renal outcomes of concern. Variables with skewed distributions were log transformed to normality before statistical tests were carried out, and tables reporting univariate comparisons give the geometric mean, SD, and range for both exposed and non-exposed groups and the p value for the statistical test. Table 4 presents two of the results, showing evidence of increased tubular and glomerular dysfunction.

Case-control studies
A case-control study begins with the identification of people with a particular condition of interest (cases) and a suitable reference group without the condition (controls). The frequency of a risk factor of interest—for example, exposure—is compared between those with the condition and those without. The success of a case-control study depends on the identification of all available cases within a given population and period and the unbiased selection of controls, so it is essential that the procedure for doing this is fully explained. The criteria for defining a case and the ascertainment of all possible cases should be made explicit.

Controls must be selected so as to be representative of those who, had they developed the condition, would have been selected as cases. Authors should describe how they decided on the numbers of controls per case, the sampling frame from which the controls were chosen, the selection method, whether controls were matched, and if so, on which variables, and the

Table 4 Comparisons of workers exposed to cadmium and hospital controls from the cross sectional study reported by Thun et al (1989)[23] investigating kidney dysfunction in workers at a cadmium production plant

Outcome	Exposed to Cd (n=45)			Hospital controls (n=32)			p Value
	Geometric mean	(SD)	Range	Geometric mean	(SD)	Range	
β-2-microglobulin in urine (µg/g creatinine)	470	(4.4)	98–107143	190	(1.6)	81–565	0.0001
Creatinine, serum (µg/d*l*)	1.16	(1.3)	0.7–2.6	1.01	(1.2)	(0.7–1.4)	0.006

reasons for the choice of variables. In most studies there are potential confounding factors and the reasons for including or excluding these should be discussed.

Analyses of case-control studies usually follow a logical sequence from calculation of ORs for levels of the risk factor, through stratified analysis to control for confounders, to modelling. As reported earlier a common error is to ignore the matching in a matched case-control study, and in particular, to carry out logistic (unconditional) rather than conditional logistic regression analysis. This can lead to an erroneous estimate of the OR. The use of an unmatched analysis of data collected in a matched design results in a bias which tends towards the null. Relative risk estimates from unmatched analyses tend, on average, to be closer to unity than those calculated from the matched sets.[17] Schlesselman[20] also points out that if matching of variables which are confounders is ignored, estimates of ORs are again biased towards unity.

The independent variables can be analysed as either continuous variables or in categories. Logistic regression models assume an exponential relation between disease risk and other continuous variables such as exposure. Rothman[24] points out that this assumption may not always be appropriate and categorising facilitates estimation of ORs for different levels of exposure without constraint to any specific pattern.

The strategy for choosing the categories should be given. For example, this could be the use of preset or standard cut off values, through examination of the distribution before disclosing case or control status, or the use of equally spaced categories—such as quartiles or quintiles.

An example of a clearly reported case-control study is that by Schnatter *et al*[25] in *Occupational and Environmental Medicine*. This was a nested case-control study investigating lymphohaematopoietic malignancies and exposure to benzene in Canadian petroleum distribution workers. The authors give a detailed description of the identification of cases, selection of matched controls, (four per case) exposure assessment method, and collection of data on potential confounders.

In their description of the statistical analysis they define the dependent variable, the primary and other independent variables, the confounders included, and the reasons for excluding others. They were concerned about a "cut off point effect" in categorising cumulative benzene exposure and therefore explored the results from several schemes derived from the distribution of exposure in the controls—for example quartiles, tertiles, and category mid points corresponding to regulatory standards.

Conditional ORs (which they also define) from univariate analyses were calculated with the Mantel-Haenzel technique, and logistic regression models were used to examine the effects of potential confounders. The authors give the computer package used, in this case EGRET. They present results firstly for the various exposure metrics used and for the confounding variables. Table 5 gives some selected univariate results. In describing these, the authors comment on the strongest risk factors, limitations due to, for example, missing data, as occurred for smoking, and the width of some of the confidence intervals. They also interpret the values of the ORs. For example, they thought that the results for the highest categories of cumulative exposure to benzene (whatever the cut off point scheme chosen) suggest risks consistent with unity.

Table 5   Selected results from a case-control study of lymphohaematopoietic malignancies and exposure to benzene in Canadian petroleum distribution workers (Schnatter et al 1996)[25]

	Exposed cases (n)	OR (95% CI)
Potential confounders:		
Socioeconomic job type:		
Managerial or professional	4	1.00
Clerk or technician	4	0.34 (0.03 to 3.10)
Operator or driver	6	0.41 (0.07 to 2.33)
Smoking status:		
Never	0	1.00
Ever	7	∞
Family cancer:		
No	8	1.00
Yes	5	2.51 (0.51 to 13.3)
Chest *x* ray film (n):		
0–9	7	1.00
10–14	3	1.41 (0.18 to 11.2)
15–19	2	0.75 (0.01 to 18.8)
⩾20	1	1.73 (0.02 to 156)
Cumulative exposure:		
Benzene (by quartiles):		
0.0–0.17	2	1.00
0.18–0.49	8	5.06 (0.34 to 295)
0.50–7.9	1	0.88 (0.01 to 18.2)
8.0–219.8	3	2.11 (0.10 to 138)
Benzene (by tertiles):		
0.0–0.22	3	1.00
0.23–5.49	8	4.37 (0.72 to 48.6)
5.50–219.8	3	0.92 (0.10 to 11.2)
Benzene (by median, 75th and 90th percentiles):		
0.0–0.49	10	1.00
0.50–7.99	1	0.22 (0.0 to 1.82)
8.0–19.99	1	0.42 (0.01 to 3.95)
20.0–219.8	2	0.96 (0.09 to 6.81)
Benzene (by regulatory standards):		
0.0–0.45	10	1.00
>0.45–4.5	1	0.43 (0.01 to 4.05)
>4.5–45	1	0.16 (0.0 to 1.32)
>45	2	1.47 (0.16 to 13.1)
Benzene (by regulatory standards):		
0.0–0.90	10	1.00
>0.90–9.9	2	0.43 (0.04 to 2.36)
>9.9–99.9	1	0.48 (0.01 to 4.55)
>99.9	1	1.03 (0.02 to 20.3)

Cohort studies

In a cohort study a group (cohort) of subjects is identified and then followed up to ascertain incidence of a particular condition. The risk of the condition is estimated in those exposed to a certain risk factor relative to those not exposed. As with cross sectional and case-control studies it is essential to describe how the study population was defined. In the occupational setting, cohorts are often historical—that is, enumerated as of some earlier time and followed up to the present. Checkoway *et al*[21] also distinguishes between a fixed cohort, in which the cohort is restricted to subjects employed or hired on some given date, and a dynamic cohort which also includes workers hired subsequently. As well as the usual considerations of sample size, the duration of follow up needs to be taken into account to allow sufficient time between exposure and occurrence of the conditions of interest. This is particularly

important for a rare disease and diseases—such as chronic disease and cancers—which may have a long latent period.

The identification of an occupational cohort is often made from personnel records, with other data sources, such as medical or union records, being used as ancillary data sources. However, it is the quality of the data which is of real interest to the reader, not necessarily the source, and comment on the accuracy and completeness of the information is desirable.

The most common health outcome of occupational cohort studies reported in *Occupational and Environmental Medicine* is mortality, with cancer incidence less often described. In the United Kingdom and Nordic countries obtaining this information is fairly straightforward, and follow up rates are usually high. However, in many other European countries, and in the United States the process may be more cumbersome and the potential problems and resulting biases should be addressed in papers. Inaccuracies and incompleteness of health outcome data should also be discussed where appropriate.

The usual measurement of risk is either the relative risk or some standardised rate—such as the standardised mortality ratio (SMR)—when stratification for such variables as age and calendar period are taken into account. Measures such as the SMR have shortcomings which are rarely considered by authors in *Occupational and Environmental Medicine*. For example, any summary measure such as an SMR obscures stratum specific effects, and it may be relevant to evaluate which stratum—such as an age group—experiences the greatest relative disease excesses or deficits. In many cohort studies analyses are also carried out separately for different exposure subcohorts, defined, for example, by job or task. Comparison of the summary measure between these subcohorts may not be appropriate if they differ in their distributions of a particular confounding factor such as age. This problem can be overcome by calculating standardised risk ratios. The use of an internal reference group may reduce bias due to the healthy worker effect. The assignment of person-years in subcohorts defined by jobs and for analyses of

disease induction and latency are discussed in detail by Checkoway *et al*.[21]

A cohort study by Sorahan *et al*[26] that reported results from indirect standardisation with an external comparison population (England and Wales) and from Poisson regression modelling is another example of a study of exposure to cadmium, in this case cadmium alloy workers and workers employed in the vicinity of copper cadmium alloy work. The authors describe in detail how the study population was defined, the reasons for any exclusions, the collection of information on work history and where this was incomplete, and the tracing of the cohort for vital status. They also describe the process used in the factories for copper cadmium alloy production and how the cumulative exposure to cadmium was estimated. In presenting their results from the indirect standardisation the authors report the cause of death, the international classification of disease (ICD) code, the observed and expected deaths, the SMR, and 95% confidence intervals (95% CIs). Table 6 gives the results for lung cancer and non-malignant diseases of the respiratory system for different occupational groups and time from the start of employment. Lung cancer shows a significant excess of observed deaths compared with those expected for vicinity workers and with excesses occurring after 20 years of follow up. All three subcohorts of workers were found to have significant excesses for diseases of the respiratory system.

Presenting the results of multivariate models

The nature of the relation between risk factors and disease is often complex. The correct interpretation of study results thus depends on taking account of the effects of covariates. As Callas *et al*[27] pointed out, advances in computer software have resulted in multivariate modelling becoming a standard method for adjustment of confounders in epidemiological research.

In cohort studies three types of regression analysis are commonly used—namely, proportional hazards, Poisson, and logistic. Proportional hazards modelling is generally consid-

Table 6   Mortality from lung cancer and non-malignant diseases of the respiratory systems, by time from starting employment, in subgroups of cadmium workers, 1946-92 (Sorahan et al 1995)[26]

Time from starting employment	Alloy workers			Vicinity workers			Brass and iron foundry workers		
	Obs	Exp	SMR	Obs	Exp	SMR	Obs	Exp	SMR
Cancer of lung†:									
1–9	1	0.6	182	5	2.2	230	2	1.0	210
10–19	4	2.4	167	5	6.2	80	1	2.8	35
20–29	4	4.5	90	21	10.4	202**	5	5.1	99
30–39	4	5.4	74	19	10.4	183*	6	4.9	122
≥40	5	5.1	99	5	5.1	97	5	4.1	122
Total	18	17.8	101	55	34.3	160**	19	17.8	107
Diseases of respiratory system (non-malignant)‡:									
1–9	2	0.7	282	4	3.0	133	1	1.1	93
10–19	7	3.2	218	13	7.7	168	5	2.5	203
20–29	17	5.6	301**	25	12.7	196**	7	4.6	152
30–39	13	3.6	206**	16	12.7	126	16	4.7	339***
≥40	15	7.7	196**	13	6.8	191*	5	4.3	118
Total	54	23.5	230***	71	43.0	165***	34	17.1	199**

*p<0.05; **p<0.01; ***p<0.0001.
†ICD 8th: 162–163.
‡ICD 8th: 460–519 and 866.

Table 7   Best fitting regression models for renal outcomes from Thun et al (1989)[23]

	β	SE	F	p Value	R^2
β-2-microglobulin (log μg/g creatinine):					
Intercept	2.323	—	—	—	
Dose ($\times 10^{-4}$)	4.195	0.498	71.1	<0.0001	0.50
Retinol binding protein (log μg/g creatinine):					
Intercept	1.457				
Dose ($\times 10^{-4}$)	3.657	0.6633	55.6	<0.0001	0.45
Age (y)	0.010	0.005	4.2	0.044	0.03
Serum creatinine (log mg/dl):					
Intercept	0.008				
Dose (x 10^{-4})	0.414	0.094	21.9	0.0001	0.23
Prostatic disease	0.050	0.029	2.9	0.09	0.03

ered to be the most appropriate method for cohort studies.[21] However, Poisson regression, which may be less costly in terms of computer time, generally gives similar results.[28] In an analysis of 200 occupational cohort studies published in 1990–1, Callas *et al*[27] found 14 studies which used an external reference group and carried out Poisson regression. A further 40 used internal comparisons with use of proportional hazards, Poisson, and logistic regression divided about equally between these studies. Callas *et al* draw attention to the fact that logistic regression does not account for possible different durations of follow up for cohort members or for changes in values of variables over time.

By contrast with cohort studies, logistic regression is the method of choice in case-control studies (conditional if matching is used). Thompson[29] gives a comprehensive review of the statistical analysis of case-control studies, including discussion of the approaches to categorisation of continuous exposure variables and the assessment of a dose-response relation.

After examination of univariate associations between disease and exposure, modelling can be used to investigate the effect of confounders on the disease-exposure relation—that is, whether inclusion affects the magnitude and direction of the relation—to assess the significance of the exposure variable in the presence of confounders, and to examine the interaction of confounders with the exposure variable. It is essential that authors state clearly which type of model they are using, and explain the assumptions and justification for their choice. Analyses of this kind may involve the fitting and comparison of many different models. In presenting the results, it is useful to give the variables included in each model, the degrees of freedom, the deviance (or other goodness of fit statistic), and the hypothesis under consideration. Comparison of pairs of models should

be made explicit to the reader, including the differences between deviances and the results of statistical testing. If the actual model equations are presented then an interpretation of the model coefficients should be given—for example, converting them to the risk estimates, such as ORs, explaining what they mean in terms of risk, and discussing their relation with other variables. It may also be appropriate to show both adjusted and unadjusted risk estimates, with of course, their CIs. Finally, some discussion of the overall goodness of fit of models is required to assess the amount of variation accounted for.

The illustrative examples given for the reporting of cross sectional, case-control, and cohort studies also report the results from modelling of their data. In the study by Thun *et al* of exposure to cadmium and kidney function multiple linear regression was used to investigate which variables best explained the renal tubular and glomerular outcomes. Explanatory variables included age, blood pressure, ethnicity, months since last exposure to cadmium, Quartelet index, body surface area, history of hypertension, prostatic disease, diabetes, and blood lead concentration. Table 7 presents the best fitting models for selected outcomes, and for each model gives the regression coefficients, their standard error, the *F* statistic, and the p value for each variable, and a measurement of the goodness of fit (R^2). Dose (cumulative exposure) was the single most important variable associated with all of the renal outcomes. The authors suggest that the positive associations (as seen from the regression coefficients) of dose with β-2-microglobulin and retinol binding protein (measures of tubular proteinuria) are consistent with a renal tubular toxin that impairs reabsorption of these substances and that the positive association of dose with serum creatinine is consistent with a glomerular effect.

Schnatter *et al* report results from conditional logistic models, which explore the influence of confounding variables on the risk estimates for exposure to benzene. For each model they present the variables included, the ORs and CIs for each variable, the significance of each variable in the model and the p value—that is, the overall goodness of fit of the model. Table 8 gives some selected results from the paper and shows the changes in the ORs as different variables are gradually added into the model. The authors are careful to point out limitations with and caveats from these analyses. For example, as the analyses only include cases and controls for which all three items of data were available—that is, a different set of cases and controls—the results are not directly comparable with others presented in the paper. They suggest that a family history of cancer and cigarette smoking may be relevant risk factors for leukaemia in the petroleum distribution workers but again, urge caution due to the missing data.

In the study of Sorahan *et al* of chromium workers Poisson regression modelling was used to adjust, simultaneously, the risk estimates for cancer of the lung and non-malignant diseases of the respiratory system with cadmium

Table 8   Selected conditional logistic modelling results for leukaemia for cases and controls with known values for potential confounders, from Schnatter et al (1996)[25]

Model p value*	Variable	OR (95% CI)	Variable p value†
0.11	Family history of cancer	5.53 (0.54 to 56.6)	0.15
0.02	Family history of cancer	14.0 (1.04 to 188.0)	0.05
	Ever smoked cigarettes	8.89 (0.66 to 119.0)	0.10
0.06	Family history of cancer	11.5 (0.83 to 160.0)	0.07
	Ever smoked cigarettes	6.93 (0.48 to 100.0)	0.16
	Cumulative exposure to benzene	0.97 (0.82 to 1.15)	0.72

*Based on score statistic.
† Based on Wald χ^2 statistic.

Table 9　Relative risks in copper cadmium alloy workers for cancer of the lung and chronic disease of the respiratory system by level of cumulative exposure, from Poisson regression modelling (Sorahan et al 1995)[26]

Cumulative exposure to cadmium ($\mu g.m^{-3}.y$)	Deaths n	RR‡ (95% CI)	Likelihood ratio test§ p value	RR ¶ (95% CI)	Likelihood ratio test§ p value
Cancer of lung††:					
<1600	10	1.0		1.0	
1600–	5	0.50 (0.17 to 1.46)		0.85 (0.27 to 2.68)	
≥4800	4	0.46 (0.14 to 1.49)	0.289	0.81 (0.18 to 3.73)	0.947
Evaluation of trend‡‡		0.65 (0.36 to 1.18)	0.144	0.89 (0.43 to 1.84)	0.753
Chronic diseases of respiratory system (non-malignant)§§:					
<1600	8	1.0		1.0	
1600–	28	3.46** (1.58 to 7.59)		4.54*** (1.96 to 10.51)	
≥4800	24	3.32** (1.48 to 7.42)	0.001	4.74** (1.81 to 12.43)	<0.001
Evaluation of trend		1.61** (1.15 to 2.23)	0.004	1.96** (1.27 to 3.02)	0.002

*P<0.05; ** P<0.01; *** P<0.001.
†Based on the mortality experience of the 347 copper cadmium alloy workers.
‡Adjustment for age.
§Difference between the deviance for a model that includes the variable cumulative exposure to cadmium and the deviance for a model that excludes this variable. This difference is interpreted as a χ^2 value with degrees of freedom equal to the number of terms added to the fuller model (three terms for the analyses by level of exposure, or one term for the analyses which evaluated trends).
¶Adjustment for age, year of start of alloy work, factory, and time since starting alloy work.
††ICD-8, 162-163, underlying or contributory cause.
‡‡Relative risk for change in exposure of one level.
§§ICD-8, 490-519, underlying or contributory cause.

exposure for age, year of starting alloy work, factory, and time from starting alloy work. The categories (levels) for these variables are specified in the paper. The authors present the number of deaths, the risk estimates (relative risk), the 95% CI, the likelihood ratio test p values for each model, and an evaluation of trend. Table 9 presents selected results. The inclusion of the variable cumulative exposure to cadmium made a highly significant improvement to the models for non-malignant diseases of the respiratory system. In discussing their results, Sorahan *et al* comment on the potential for some of the findings, particularly between factories, to be an artefact in data collection, the reliability of the exposure estimates, and the limitations of the data—for example, a lack of smoking history information.

Discussion

In 1981 some guidelines for the documentation of observational epidemiological studies were published in the *American Journal of Epidemiology*.[30] They were the culmination of lengthy discussion in the American academic community, public sector, business organisations, and United States regulatory agencies concerned with public health. The aim of the documentation guidelines was to assist the regulatory agencies in the "objective and scientific evaluation of epidemiological studies, as they bear on public health decisions". Although they do not refer specifically to the publication of epidemiological research in scientific journals, they provide a concise and structured description of the important elements of a study which should be documented for the (*a*) background and objectives, (*b*) study design, (*c*) study and comparison subjects, (*d*) data collection procedures, (*e*) analysis, and (*f*) supporting documentation.

Since then there has been increasing interest, particularly in the United States, in the development and application of more formal "good epidemiology practices". The Chemical Manufacturers Association published guidelines for good epidemiology practices for occupational and environmental epidemiological research in 1991.[31] The non-experimental nature of occu-

pational and environmental epidemiological studies can, as readers of *Occupational and Environmental Medicine* will be aware, spark controversy, particularly concerning the interpretation and significance of epidemiological study results. The Chemical Manufacturers Association guidelines were developed to consider the issues of data quality, study design, and study conduct which are under the control of the investigator, and to improve confidence in the use of epidemiology in the formulation of public health policy. They give detailed requirements concerning research personnel and facilities, development of the study protocol, review and approval, study conduct (including protection of human subjects, data collection, verification, analysis, and study reporting), communication, archiving, and quality assurance. For many of the guideline requirements they advocate the use of standard operating procedures, which are written detailed descriptions of routine procedures involved in performing epidemiological studies, for example, collecting raw data, coding death certificates, etc. They also emphasise the need for constant review, scientific, ethical, and administrative, at all stages of a study.

A major implication of applying more formal good epidemiology practices to research is the potential for costs and time to complete the work to be increased. Development and application of standard operating procedures, for example, will inevitably add to these. In most studies carried out by the academic community many of the suggestions in the Chemical Manufacturers Association guidelines are applied in an informal manner. However, if occupational and environmental epidemiology is to succeed in achieving its goal of contributing valuable research to pubic health, perhaps it is time that those initiating the research considered the wider use of good epidemiology practices. More rigorous application of these methods should have the beneficial effect of leading, not only to better designed and executed studies, but to more rigorous reporting of the results in the scientific literature.

1 Altman DG. Statistics and ethics in medical research: misuse of statistics is unethical. *BMJ* 1980;**281**:1267–9.

2 Altman DG, Bland JM. Presentation of numerical data. *BMJ* 1996;**312**:572.
3 Altman DG, Bland JM. Absence of evidence is not evidence of absence. *BMJ* 1995;**311**:485.
4 Emmerson JD, Colditz GA. Use of statistical analysis in the New England Journal of Medicine. *N Engl J Med* 1983;**309**:709–13.
5 Reznick RK, Dawson-Saunders E, Folse JR. A rationale for the teaching of statistics to surgical residents. *Surgery* 1987;**101**:611–17.
6 Altman DG. Statistics in medical journals: developments in the 1980s. *Stat Med* 1991;**10**:1897–913.
7 Gore SM, Jones IG, Rytter EC. Misuse of statistical methods: critical assessment of articles in the BMJ from January to March 1976. *BMJ* 1977;**1**:85–7.
8 Pocock SJ, Hughes MD, Lee RJ. Statistical problems in the reporting of clinical trials. A survey of three medical journals. *N Engl J Med* 1987;**317**:426–32.
9 MacArthur RD, Jackson GG. An evaluation of the use of statistical methodology in the Journal of Infectious Diseases. *J Infect Dis* 1984;**149**:349–54.
10 McGuigan SM. The use of statistics in the British Journal of Psychiatry. *Br J Psychiatry* 1995;**167**:683–8.
11 White SJ. Statistical errors in papers in the British Journal of Psychiatry. *Br J Psychiatry* 1979;**135**:336–42.
12 McNamee R, Cockcoft A. Statistical and epidemiological reviewing. *Occup Environ Med* 1994;**51**:721.
13 Altman DG. *Practical statistics for medical research*. London: Chapman and Hall, 1991.
14 Nemery B, Casier P, Roosels D, *et al*. Survey of cobalt exposure and respiratory health in diamond polishers. *Am Rev Respir Dis* 1992;**145**:610–16.
15 Campbell MJ, Machin D. *Medical statistics a common sense approach*. Chichester: John Wiley, 1993.
16 Gardner MJ, Altman DG. *Statistics with confidence*. London: British Medical Journal Publishing Group, 1989.
17 Breslow NE, Day NE. *Statistical methods in cancer research*. Vol I. *The analysis of case-control studies*. Lyon: International Agency for Research on Cancer, 1980. (IARC Sci Publ No 32.)
18 Breslow NE, Day NE. *Statistical methods in cancer research*. Vol 2. *The design and analysis of cohort studies*. Lyon: International Agency for Research on Cancer, 1987. (IARC Sci Publ No 82.)
19 Kahn HA, Sempos CT. *Statistical methods in epidemiology*. Oxford: Oxford University Press, 1989.
20 Schlesselman JJ. *Case-control studies. Design, conduct, analysis*. Oxford: Oxford University Press, 1982.
21 Checkoway H, Pearce N, Crawford-Brown DJ. *Research methods in occupational epidemiology*. Oxford: Oxford University Press, 1989.
22 Copeland KT, Checkoway H, McMichael AJ, *et al*. Bias due to misclassification in the estimation of relative risk. *Am J Epidemiol* 1977;**105**:488–95.
23 Thun MJ, Osorio AM, Schober S, *et al*. Nephropathy in cadmium workers: assessment of risk from airborne occupational exposure to cadmium. *Br J Ind Med* 1989;**46**:689–97.
24 Rothman KJ. *Modern epidemiology*. Boston: Little, Brown, 1986.
25 Schnatter AR, Armstrong TW, Nicholich MJ, *et al*. Lymphohaematopoietic malignancies and quantitative estimates of exposure to benzene in Canadian petroleum distribution workers. *Occup Environ Med* 1996;**53**:773–81.
26 Sorahan T, Lister A, Gilthorpe MS, *et al*. Mortality of copper cadmium alloy workers with special reference to lung cancer and non-malignant diseases of the respiratory system. *Occup Environ Med* 1995;**52**:804–12.
27 Callas PW, Pastides H, Hosmer DW. Survey of methods and statistical models used in the analysis of occupational cohort studies. *Occup Environ Med* 1994;**51**:649–55.
28 Clayton D. The analysis of event history data: a review of progress and outstanding problems. *Stat Med* 1988;**7**:819–41.
29 Thompson DW. Statistical analysis of case-control studies. *Epidemiol Rev* 1994;**16**:33–50.
30 Epidemiology Work Group of the Interagency Regulatory Liaison Group. Guidelines for documentation of epidemiologic studies. *Am J Epidemiol* 1981;**1114**:609–18.
31 The Chemical Manufacturer's Association. *Guidelines for good epidemiology practices for occupational and environmental epidemiologic research*. Washington: Chemical Manufacturers Associaton, 1991:1–25.

Correspondence and editorials

Occupational and Environmental Medicine welcomes correspondence relating to any of the material appearing in the journal. Results from preliminary or small scale studies may also be published in the correspondence column if this seems appropriate. Letters should be not more than 500 words in length and contain a minimum of references. Tables and figures should be kept to an absolute minimum. Letters are accepted on the understanding that they be subject to editorial revision and shortening.

The journal also publishes editorials which are normally specially commissioned. The Editor welcomes suggestions regarding suitable topics; those wishing to submit an editorial, however, should do so only after discussion with the Editor.