SELECTING APPROPRIATE STUDY DESIGNS TO ADDRESS SPECIFIC RESEARCH QUESTIONS IN OCCUPATIONAL EPIDEMIOLOGY

Harvey Checkoway, Neil Pearce, David Kriebel

Historically, occupational epidemiology studies have often been initiated in response to concerns about apparent workplace hazards. Such concerns typically are motivated by observations of disease clusters in a workforce, findings from previous epidemiological studies of similar workplace settings, or evidence derived from other disciplines, such as toxicology, suggesting potential health impairment from workplace exposures.

Occupational epidemiologists employ a variety of study approaches to investigate work-related illness and injuries. Many of these are familiar designs that are commonly applied in other branches of epidemiology, but some are characteristic to occupational studies. The choice of study design is nearly always determined by the research question of interest, and by feasibility constraints. In this brief review, we will summarise some study design features, including particular strengths and limitations, with an emphasis on selecting the design that should be most appropriate for investigating the exposure/health outcome association of interest. Readers seeking more in-depth discussions of study designs are encouraged to consult text books on occupational epidemiology[1–3] or general epidemiology texts.[4 5]

► BASIC PRINCIPLES OF STUDY DESIGN

A fundamental concept that underpins all epidemiological research is the requirement for clearly defining the *source population*, also known as the *study base*.[6] In studies of occupational risk factors for disease and injuries, the source population should be a cohort of workers from one or more industries. Identifying the source population is relatively straightforward when conducting a study of a well-defined cohort of workers from a particular industry or facility, as is typical of most occupational cohort mortality studies. Less well appreciated is that a study that focuses on a certain health outcome, and seeks to identify multiple possible occupational risk factors, such as a population-based case-control study, has an implicit source population that generated the cases, namely the general population that includes workers from the industries and occupations of interest, workers from other industries, and non-employed persons. For example, consider a community-based case-control study of occupational risk factors for Parkinson's disease in which associations are estimated for employment in various occupations, such as farming, welding, and teaching, as well as associations with certain exposures that may span numerous occupations, such as pesticides, metals and infectious agents. In this situation, the source population would include a number of different subpopulations defined by occupation (farmers, welders, teachers) or by exposure (pesticides, metals, infectious agents). An underlying validity principle is that the controls' exposures in the case-control study should represent the exposure experience of the source population.

A second important point is that the new occurrence of disease, *incidence*, is the basic measure of disease occurrence that epidemiologists seek to estimate. Measuring new onset of illness or injury is largely unambiguous for acute health outcomes, such as non-fatal workplace injuries. Mortality is a special type of incidence in which the "event" is death rather than the occurrence of (non-fatal) disease or injury. It therefore is often used as a surrogate for disease incidence for diseases that are usually fatal (for example, cancer), but may also be affected by factors that affect survival as well as risk factors for disease incidence. Determining disease incidence is especially challenging for conditions that do not have sharp times of onset, even when serial health measurements are made. Coal worker's pneumoconiosis is a disease that fits this description. Many health outcomes develop over prolonged time periods in which onset times can only be inferred from indirect evidence. This is the case for conditions such as chronic obstructive lung disease, but is also true for diseases such as cancer for which there is generally a single diagnostic point in time, but the underlying disease

See end of article for authors' affiliations

Correspondence to:
Professor H Checkoway, University of Washington, Department of Environmental and Occupational Health Sciences, Box 357234, Seattle, WA 98195, USA; checko@u.washington.edu

process may have developed over many years. It should also be appreciated that chronic disease onset times are typically classified as single events, such as dates of disease diagnosis or death, although true disease onset is a continuous phenomenon that is difficult to characterise epidemiologically.

In certain situations (for example, cognitive impairment), determining the onset of incident disease may be impractical, and thus disease prevalence is studied instead. Although disease prevalence may be a surrogate for incidence, it is also affected by factors that determine the duration of disease (including factors that affect survival or treatment efficacy) in addition to risk factors for disease incidence. This is not to say that studies based on prevalence are inherently flawed or invalid, although distinguishing associations of health outcomes with occupational exposures that pertain to disease aetiology from those that may be related to disease severity, prognosis and duration can be difficult, if not impossible, when prevalent cases are included in a study.

STUDY DESIGN OPTIONS

As we will review below, each study design option has various features that make it more or less suitable for investigating particular exposure/disease relations. A summary of the types of health outcomes and the corresponding study design choices is shown in table 1. It should be appreciated that some research questions can be investigated by more than one epidemiological approach, but one design is usually clearly preferable for providing direct causal evidence.

CONVENTIONAL STUDY DESIGNS
Cohort studies

The cohort design entails follow-up of a population and determination of the subsequent incidence of health outcomes. Cohort studies can be classified according to their temporal sequence, either historical (retrospective) or prospective. Prospective cohort studies are particularly well suited for investigations of relatively short-term phenomena, such as pregnancy outcomes, in which the temporal relation between exposure and subsequent risk is relatively short. The span of a prospective cohort study may be as short as a single work shift (for example, across-shift lung function change), a work-week (for example, exacerbation of symptoms), or may extend to years or decades (for example, incidence of injuries). The logistical difficulties of performing prospective cohort studies, especially following study subjects and updating exposure data,

over many years, represents a serious feasibility constraint. The historical cohort design was originally developed as a more practical alternative to prospective studies for investigating diseases with long induction and latency periods, and has since become the mainstay of occupational studies of mortality and incidence from chronic diseases. Typically, historical cohort studies are limited to mortality outcomes because, unlike data for non-fatal outcomes, mortality data are readily available in most countries. An exception would be an industry that maintains a health surveillance database that would accommodate investigations of non-fatal conditions. A common prominent limitation of historical cohort studies is absent or sparse data on past exposures.

The cohort design has an intuitive logical appeal in that the temporal sequence from exposure to disease outcome mimics the widely recognised approach of an experimental paradigm, such as a randomised clinical trial. Nonetheless, temporality of exposure and outcome can also be determined validly with other study designs.

Cross-sectional studies

The cross-sectional design involves comparisons of disease prevalence among exposed and non-exposed groups, or among groups classified according to exposure type and level. Subject selection is usually based on exposure status. It is also possible to select subjects on the basis of health status, but in this situation the study is really a case-control study of prevalent health conditions rather than a standard cross-sectional study (which would include exposed and non-exposed subjects irrespective of their health status).

Cross-sectional studies are most appropriate for studying relatively persistent conditions, rather than transient or reversible effects of exposure. Typical health outcomes investigated with the cross-sectional design are repetitive motion musculoskeletal disorders, chronic respiratory impairment, and the pneumoconioses. In addition, physiological abnormalities, such as diminished lung function or elevated liver enzymes, and indicators of biological damage at the cellular level, such as chromosome abnormalities, are also amenable to study with the cross-sectional design.

Cross-sectional studies are often criticised for providing limited causal inference because exposure and health outcomes are usually assessed concurrently. In other words, such studies may be prone to a ''reverse causation'' bias—that is, the exposure status may be an effect of the disease rather than a

Table 1 Design options for studies of occupational exposures and categories of health outcomes

Health outcome	Example	Study design options*
Chronic disease (rare)	Stomach cancer	1. Historical cohort (incl. nested case-control) 2. Community-based case-control 3. Prospective cohort
Chronic disease (not rare)	Asthma onset (not asthma attack or exacerbation)	1. Prospective (inception) cohort 2. Case-control 3. Historical cohort
Acute fatal	Injury	1. Historical cohort 2. Case-crossover 3. Nested case-control
Acute non-fatal	Neurological symptoms	1. Case-crossover 2. Cross-sectional

*Listed in descending priority order.

cause. This could occur, for example, if a worker changed departments (from a more dusty to a less dusty job) or left employment as a result of developing respiratory disease. This shortcoming is not an inherent flaw of the cross-sectional design, especially in situations where a full accounting of exposure history (rather than merely current exposure status) is ascertained, as was done in the study by Eisen et al[7] of asthma among US automotive workers exposed to metalworking fluids. Nonetheless, the cross-sectional design may be particularly prone to the healthy worker survivor effect[8] in situations where only actively employed workers are studied. This form of bias may lead to missed or underestimated associations if the most heavily exposed, and consequently the most severely affected workers, have preferentially left employment and are hence not available for study. Attempts to identify and include former workers, although logistically challenging, can mitigate this bias.

Repeated measures studies

There are alternatives to the cross-sectional design to examine non-fatal health endpoints or physiological damage indicators. The best developed of these is the repeated measures study in which exposures and health status are determined at a baseline time point, and re-assessed throughout a period of follow-up. Relatively short-term follow-up (for example, several years) can provide the framework for longer-term investigations of chronic effects, such as myocardial infarction and stroke. Repeated measures studies share the identical design as prospective cohort studies. The distinction between the two is generally the nature of the health outcomes studied with these approaches: disease incidence or mortality in prospective cohort studies; and disease symptoms and physiological parameter changes in repeated measures studies.

The optimal study populations for follow-up are inception cohorts of newly hired, and hence newly exposed, workers. A good template for this approach is the 20-year prospective follow-up of respiratory system outcomes among Chinese cotton textile workers conducted by Christiani et al.[9–11] Findings from the first five years of follow-up demonstrated accelerated loss of lung function,[9] and subsequent follow-up findings indicate the potential for chronic obstructive lung disease related to cotton dust and endotoxin exposure.[10 11] Inclusion in this study required workers to have had a minimum of two years' employment to ensure follow-up of a stable workforce; thus, this was not strictly an inception cohort of new hires. Assembling cohorts of new hires, although desirable from the standpoint of investigating new onset disease in relation to initial and subsequent exposure, can pose logistical difficulties. Enrolment may suffer from high turnover rates in the early weeks or months of employment, and accumulating sufficiently large numbers of new hires in industries with sporadic hiring practices may require prolonged recruitment periods.

Case-control studies

Case-control designs entail exposure comparisons made between an index case group and a reference group of persons free of the disease of interest at the times of cases' diagnoses. Typically, efforts are made to enrol all possible cases who meet study inclusion criteria, and controls are then selected as a sample of the source population that generated the cases. Case-control studies may be nested within defined occupational cohorts, or may be conducted in the community at-large (community-based studies). Both cases and controls in nested industry-based case-control studies are from the same cohort, defined variously as members of a particular facility, occupation, industry or profession. In contrast, community-based case-control studies involve multiple occupational subpopulations from the population at large. In both types of study, cases may be identified from various sources, such as hospitals, disease registers, and death or birth certificates. However, in nested case-control studies, cases may also be identified directly by a survey or surveillance of the cohort.

It has long been recognised that the case-control design has decided advantages in terms of efficiency, relative to full cohort studies. For studies of ''rare diseases'' (for example, most cancers) case-control studies offer a cost- and time-efficient means of accruing relatively large numbers of cases, thus avoiding prolonged follow-up of large cohorts. Also, the reduced study size of a case-control study, compared to a full cohort study, can permit efficient resource allocation to refining exposure assessment and obtaining data on potential confounding factors (for example, smoking) which may not be practical in a cohort study.

Control selection for nested case-control studies is a relatively straightforward matter in most instances, whereas the choice of controls in community-based studies is often more complicated and subject to uncertainty. In community-based studies, controls should be a random sample of the source population, but this may not always be well-defined or enumerated. Ideally, controls should be selected from population registers, but when these are not available, controls may be selected from other sources, such as patients in the same hospital but admitted for an illness unrelated to the exposure, neighbours or family members. There can be several alternative choices for controls for a given study, each with characteristic advantages and limitations in terms of validity, efficiency for addressing study questions of interest, and feasibility. When selecting controls, the underlying methodological principle required to maintain study validity is to select controls such that they represent the source population that generated the cases. The concept of ''counterfactual'' matching of cases and controls, such that controls would have been identified as cases had they developed the health outcome of interest during the period of observation of the study base, can also be invoked as a guideline for validity.[12]

In either type of case-control study, controls should be free of the outcome of interest (to the extent that can be determined) at the times of cases' diagnoses. Thus, it is possible for a subject to be selected as a control for a given case at one time, but subsequently be included as a case if he or she develops the outcome of interest subsequently. This selection method, known as ''incidence density'' sampling, allows for causal inferences to be drawn with equivalent validity in nested case-control and full cohort analyses.[13]

There are situations in which selecting more than one control group is desirable to minimise confounding and other biases. For example, consider a community-based case-control study of lung cancer in relation to exposure to dusty construction work. One control group might be a random sample of all members of

635

the community (free of lung cancer), and a second control group might be patients with other types of cancer not plausibly related to dusts—perhaps brain and reproductive system cancers, for example. The first control group would represent the general source population from which the cases arose, but might be biased because of differential recall between cancer cases and healthy controls. Comparisons of exposures between cases and the second control group would help minimise recall bias because all subjects would be cancer patients, and thus likely to be in a similar state of mind when asked to recall potentially hazardous exposures. If the results from comparisons with both control groups were similar, this might strengthen arguments for causality, although there may also be plausible reasons for discrepant findings.

VARIANTS OF THE CASE-CONTROL DESIGN

Over the past two decades, two variants of the case-control design, namely the case-cohort and case-crossover designs, have been developed that have clear efficiency and validity advantages over the conventional case-control design in some situations.

Case-cohort design

In a case-cohort study, there are multiple case groups and a common comparison group.[14] The latter is selected as a random sample representative of the source population (cohort) that generated the cases, and is termed the "reference subcohort". Case-cohort studies that are nested within defined occupational cohorts are far more common than community-based case-cohort studies, although a community source population would not preclude application of this design. The multisite cancer hospital-based case-cohort study in Montreal[15] is an example of the latter. The particular advantage of the case-cohort approach is that it permits efficient testing of associations with multiple health outcomes (case groups). In the conventional case-control approach, a control group would have to be selected for each case group, whereas the case-cohort design allows using one comparison group repeatedly. Because the reference subcohort is simply a random sample of the cohort or source population, it may contain subjects who are also in one of the case groups. Inclusion of cases in the reference subcohort will not introduce bias provided that exposures for subcohort members are truncated at the times when they develop the disease of interest in a specific analysis. For example, in a case-cohort analysis of stomach cancer in which the subcohort includes one or more subjects with stomach cancer (by virtue of random sampling), the overlapping cases would also be included in the case group, and their exposure histories would be included with the exposure experience of the reference subcohort up to the dates of their diagnoses. This is equivalent to incidence density matching.[14]

Application of the case-cohort design is illustrated by a study of occupational risk factors for various cancers among women workers in the Shanghai textile industry. The study was originated as an intervention trial of breast self-exam in a cohort of over 250 000 women workers.[16] Exposure assessments were performed for numerous textile industry chemicals and dusts, including fibre dusts, solvents and endotoxin.[17 18] Table 2 provides a summary of case-cohort comparisons for cumulative exposure to endotoxin by various lag intervals, for selected

gastrointestinal cancers.[19–22] In these analyses, each case group's exposures was compared with exposures experienced by a common reference subcohort of approximately 3200 workers. These findings indicate that the highest cumulative endotoxin exposures were associated with reduced risks for several different cancers, especially when exposures were lagged by 20 years, suggesting possible early-stage anticarcinogenic effects.

Case-crossover design

The case-crossover design was formulated to characterise risk factors for health outcomes that occur in close temporal sequence to exposure, especially for so-called "disease triggers".[23] To date, most applications of this design have been in studies of acute outcomes related to environmental air pollution; thus, methodological aspects of case-crossover studies have been developed in that context.

Typically, the outcomes of interest are acute events, such as injuries or disease symptoms with abrupt onsets. This design only includes an index case group, and involves the comparison of cases' exposures immediately before (or very close in time to) their events with exposures that occur at other "typical" times. Consequently, each case serves as his or her own individually matched control in which the index interval before the event is treated as the "case" and the reference interval representing typical exposures is the "control". The principal advantage of the case-crossover design, relative to a conventional case-control study, is that matching each case with himself or herself greatly facilitates control of potential confounders that are time invariant and possibly difficult to measure, such as genetic factors. However, potential confounders that are not time invariant, especially over the relatively short period of observation in a case-crossover study, such as recent infection status, will require control by conventional methods.

A study of risk and protective factors among for acute hand injuries[24] offers a good illustration of this approach. Cases provided details on the extent and timing of transient work factors during the 90 min preceding their injuries, and were classified as exposed if they experienced these factors at the time of the injury. Reference period exposures were estimated as averages for the month preceding the injury. As shown in table 3, working with unusual equipment or materials was strongly associated with increased risks, and glove use

Table 2 Relative risks for gastrointestinal cancers associated with highest cumulative exposures to endotoxin, by lag interval, among women textile workers in Shanghai, China

Cancer site (cases)	0-year lag		20-year lag	
	RR (95% CI)*	trend p†	RR (95% CI)*	trend p†
Oesophagus (102)	0.5 (0.2 to 1.2)	0.06	0.4 (0.2 to 0.8)	0.01
Stomach (646)	0.8 (0.6 to 1.1)	0.20	0.7 (0.5 to 0.9)	<0.001
Colon (477)	1.2 (0.8 to 1.6)	0.46	0.8 (0.6 to 1.1)	0.32
Rectum (274)	1.1 (0.7 to 1.7)	0.75	0.6 (0.4 to 1.0)	0.08
Pancreas (180)	0.8 (0.5 to 1.2)	0.12	0.5 (0.3 to 0.9)	<0.001
Liver (360)	0.8 (0.6 to 1.1)	0.25	0.6 (0.4 to 0.9)	0.02

*Relative risk (95% CI) for the highest versus lowest exposure strata, adjusted for age and cigarette smoking.
†Exposure-response trend p value.

Table 3 Relative risks for hand injury associated with transient workplace conditions

Worker group	Unusual work task or equipment	Glove use
All subjects	11.0 (9.4 to 12.8)*	0.4 (0.3 to 0.5)
Occupational group		
Machine/assembly	10.6 (18.5 to 13.3)	0.5 (0.4 to 0.6)
Construction trades	17.4 (9.3 to 32.3)	0.3 (0.2 to 0.4)
Packaging	5.1 (3.8 to 6.9)	0.4 (0.3 to 0.7)
Service, prof, mgmt	15.7 (11.1 to 22.3)	0.4 (0.3 to 0.6)
Job experience (years)		
≤1	13.2 (9.1 to 19.2)	0.4 (0.3 to 0.6)
1 to 3	16.3 (11.4 to 23.3)	0.2 (0.2 to 0.4)
>3	9.0 (7.4 to 11.0)	0.5 (0.4 to 0.6)

*Relative risk (95% CI) exposed versus non-exposed.

637

Main messages

► Various epidemiological study designs have particular strengths and limitations for investigating particular exposure/disease relations.

► Study design selection should be guided by the suitability of the design for the research question at hand, and by feasibility constraints.

► Conventional approaches, including cohort, case-control, and cross-sectional designs, should continue to be mainstay methods; application of newer variants of the case-control design—case-cohort and case-crossover studies—for specific purposes should be encouraged.

► A series of coordinated epidemiological studies whose designs are tailored to investigate specific research questions, will inevitably be required to address a wide range of occupational health concerns.

Policy implications

► Selection of the most suitable epidemiological study designs for specific research questions will be required for maximising knowledge on illness and injury risk factors, and ultimately for informing disease prevention programmes.

conferred protection. Similar findings were noted among various occupational groups and job tenures.

The selection of index and reference intervals is not necessarily clear cut, and can pose some methodological challenges. The width of the index interval will depend on the characteristics of the exposure and the health outcome and the nature of their presumed relation. In the simplest case of a very acute severe injury, the index interval can be as short as several minutes or hours, whereas for an outcome with a longer induction time (for example, myocardial infarction), the index period may be defined as one or more days. In addition, it may be necessary to include a lag interval between the index interval and the event onset time for outcomes that may be delayed manifestations of exposure. For example, the effects of sensitising chemicals may appear hours or days after relevant exposures.

The placement and width of reference intervals can be sources of uncertainty. Reference intervals are generally selected as time periods preceding index intervals, such as the preceding day, or the same day of the week during the past month.[25] Alternatively, a bi-directional control sampling scheme can be adopted such that reference intervals are selected both before and after the event occurrence times in order to control for predictable temporal changes in exposure, as might occur when air pollution levels are known to be decreasing.[26] Uni-directional sampling should be most appropriate for the majority of occupational studies for several reasons. In a workplace setting, there may be predictable changes in exposure due to changes in ventilation or use of protective equipment, although the time scale of these changes will ordinarily exceed the duration of observation of a case-crossover study. Furthermore, the bi-directional referent sampling scheme requires the assumption that case events will not influence subsequent exposures, which may hold in studies of air pollution or climatic changes, but may be violated in a workplace setting if, for example, safety enforcement policies are modified after a fatal accident occurrence.

DISCUSSION

When confronted with the task of answering questions about the relative safety of the workplace environment, occupational epidemiologists prefer to design and implement studies that allow testing of very specific exposure/disease associations. Deciding which study design is most suitable for addressing a particular occupational health question will depend on the nature of the health outcome(s) and exposure(s) of interest and, to a

great extent, on feasibility. In practice, logistical considerations frequently are the critical determinants of study design choice.

By way of illustration, consider a situation where there is concern about potential cardiovascular toxicity of a certain workplace chemical. This concern would be addressed optimally by a prospective cohort study of changes in cardiovascular disease incidence and related clinical parameters among current and former workers, whose exposures to the chemical of interest and potential confounders are assessed with a high degree of accuracy. Inclusion of newly hired workers as inception cohorts would be especially valuable for identifying early changes in health status. It becomes readily apparent, however, that the requirements of time, cost and data for such a study may far exceed available resources, thus necessitating alternative approaches. Among alternatives, cross-sectional studies may yield some aetiological insights, primarily among actively employed workers, although the likely absence of data for former and retired workers and the potential for healthy worker survival effect bias could be severe limitations. A historical cohort mortality study is another option, but could only evaluate the effects of exposure on fatal cardiovascular disease. A reasonable strategy might then be a series of epidemiological studies, each of which addresses various aspects of cardiovascular system risk. Such studies could include: targeted inception cohort studies of changes in selected cardiovascular health parameters (serum lipids, blood pressure, heart rate variability, etc); a cross-sectional study of specific health endpoints (for example, hypertension); a cohort mortality study, initiated first as a retrospective cohort study, and expanded to incorporate prospective follow-up as a component of worker health surveillance. Nested case-cohort studies of specific cardiovascular diseases in which detailed data on non-occupational risk factors are obtained and case-crossover studies to identify acute exposure-related effects would also be beneficial.

As the above hypothetical example is intended to illustrate, it is very unlikely that any single epidemiological study design can yield data adequate to investigate a broad spectrum of occupational health questions. Instead, a rational epidemiological strategy is to conduct separate, yet related studies whose designs are most suitably tailored to address specific research questions. Conventional epidemiological study designs will no doubt continue to serve as the mainstay approaches. The case-cohort and case-crossover variants of case-control studies offer distinct advantages, and their further application in occupational epidemiology should be encouraged.

ACKNOWLEDGEMENTS

. . . .

Authors' affiliations
H Checkoway, Department of Environmental and Occupational Health Sciences, University of Washington, Seattle, Washington USA and International Agency for Research on Cancer, Lyon, France
N Pearce, Centre for Public Health Research, Massey University Wellington Campus, Wellington, New Zealand
D Kriebel, Department of Work Environment, University of Massachusetts at Lowell, Lowell, Massachusetts, USA

Competing interests: None declared.

REFERENCES

1 **Monson RR**. *Occupational epidemiology*, Second edition. Boca Raton, FL: CRC Press, 1990.
2 **Hernberg S**. *Introduction to occupational epidemiology*. Chelsea, MI: Lewis Publishers, 1992.
3 **Checkoway H**, Pearce N, Kriebel D. *Research methods in occupational epidemiology*, Second edition. New York: Oxford University Press, 2004.
4 **Rothman KJ**, Greenland S. *Modern epidemiology*, Second edition. Philadelphia: Lippincott-Raven Press, 1998.
5 **Koepsell TD**, Weiss NS. *Epidemiologic methods: studying the occurrence of illness*. New York: Oxford University Press, 2003.
6 **Miettinen OS**. *Theoretical epidemiology*. NY: Wiley, 1985.
7 **Eisen EA**, Holcroft CR, Greaves IA, *et al.* A strategy to reduce the healthy worker effect in a cross-sectional study of asthma and metalworking fluids. *Am J Ind Med* 1997;**31**:671–7.
8 **Eisen EA**. The healthy worker effect in morbidity studies. *Med Lav* 1995;**86**:125–38.
9 **Christiani DC**, Ye T-T, Wegman DH, *et al.* Cotton dust exposure, across-shift drop in FEV1, and five-year change in lung function. *Am J Respir Crit Care Med* 1994;**150**:1250–5.
10 **Christiani DC**, Wang X-R, Pan L-D, *et al.* Longitudinal changes in pulmonary function and respiratory symptoms in cotton textile workers. *Am J Respir Crit Care Med* 2001;**163**:847–53.
11 **Wang XR**, Zhang HX, Sun BX, *et al.* A 20-year follow-up study of chronic respiratory effects of exposure to cotton dust. *Eur Respir J* 2005;**26**:881–6.
12 **Newman SC**. Commonalities in the classical, collapsibility, and counterfactual concepts of confounding. *J Clin Epidemiol* 2004;**57**:325–9.
13 **Lubin JH**, Gail MH. Biased selection of controls for case-control analyses of cohort studies. *Biometrics* 1984;**40**:63–75.
14 **Prentice RL**. A case-cohort design for epidemiologic cohort and disease prevention trials. *Biometrika* 1986;**73**:1–11.
15 **Siemiatycki J**. *Risk factors for cancer in the workplace*. Boca Raton, FL: CRC Press, 1991.
16 **Thomas DB**, Gao DL, Self SG, *et al.* Randomized trial for breast self-exam in Shanghai: methodology and preliminary results. *J Natl Cancer Inst* 1997;**89**:355–65.
17 **Wernli KJ**, Astrakianakis G, Camp J, *et al.* Development of a job exposure matrix (JEM) for the textile industry in Shanghai, China. *J Occup Environ Hyg* 2006;**3**:521–9.
18 **Astrakianakis G**, Seixas NS, Camp JE, *et al.* Modeling, estimation and validation of cotton dust and endotoxin exposures in Chinese textile operations. *Ann Occup Hyg* 2006;**50**:573–82.
19 **Wernli KJ**, Fitzgibbons ED, Ray RM, *et al.* Occupational risk factors for esophageal and stomach cancers among women textile workers in Shanghai, China. *Am J Epidemiol* 2006;**163**:717–25.
20 **De Roos AJ**, Ray R, Gao DL, *et al.* Colorectal cancer incidence among female textile workers in Shanghai, China: a case-cohort analysis of occupational exposures. *Cancer Causes Control* 2005;**16**:1177–88.
21 **Li W**, Ray R, Gao DL, *et al.* Occupational risk factors for pancreatic cancer among female textile workers in Shanghai, China. *Occup Environ Med* 2006;**63**:788–93.
22 **Chang C-K**, Astrakianakis G, Thomas DB, *et al.* Occupational exposures and risks of liver cancer among Shanghai female textile workers: a case-cohort study. *Int J Epidemiol* 2006;**35**:361–9.
23 **Maclure M**. The case-crossover design: a method for studying transient effects on the risk of acute events. *Am J Epidemiol* 1991;**133**:144–53.
24 **Sorock GS**, Lombardi DA, Hauser R, *et al.* A case-crossover study of transient risk factors for occupational hand injury. *Occup Environ Med* 2004;**61**:305–11.
25 **Janes H**, Sheppard L, Lumley T. Case-crossover analyses of air pollution exposure data: referent selection strategies and their implications for bias. *Epidemiology* 2005;**16**:717–26.
26 **Navidi W**. Bidirectional case-crossover designs for exposures with time trends. *Biometrics* 1998;**54**:596–605.

ECHO ....

Evidence on asthma outcome found wanting

Please visit the *Occupational and Environmental Medicine* website [www.occenvmed.com] for a link to the full text of this article.

Authors of a systematic review are advocating a large collaborative and carefully standardised study to determine best advice for patients with occupational asthma, in the light of published findings on outcome. Better evidence is needed, they say.

Patients' prospects seem poor, by the two outcomes assessed—complete symptomatic recovery and physiological recovery after antigenic exposure has ceased. About a third of patients overall achieved complete recovery from symptoms, based on 39 original studies covering 1681 patients with a median follow up of 31 (range 6–240) months. Potential factors affecting this outcome were increased age and clinic based populations, with lower pooled prevalences for 100% recovery. Physiological recovery was also limited, as non-specific bronchial hyper-responsiveness (NSBHR) persisted after diagnosis in most of the 28 studies identified, covering 695 patients with a median follow up 37(6–240) months.

However, variation among studies was significant for both outcomes and the authors suspect length of follow up may be important. Most patients were from specialist clinics—an important source of bias. The few studies of workplace populations showed a higher pooled prevalence of recovery from symptoms than clinic based populations and lower pooled prevalence of NSBHR but at best were only borderline significant.

The analysis is compromised too by variation in study size and methods, possible overlap between some study populations, and differences in the way outcomes were measured. Added snags were inability to control for other important aspects like severity of asthma at diagnosis or use of protective measures.

▲ Rachiotis G, *et al. Thorax* 2007;**62**:147–152.

638