Bypassing the selection rule in choosing controls for a case –control study

Keith T Palmer, Miranda Kim and David Coggon

Occup Environ Med 2010 67: 872-877 originally published online September 23, 2010
doi: 10.1136/oem.2009.050674

Bypassing the selection rule in choosing controls for a case–control study

Keith T Palmer, Miranda Kim, David Coggon

Community Clinical Sciences, MRC Lifecourse Epidemiology Unit, University of Southampton, Southampton, UK

Correspondence to
Professor Keith Palmer, MRC Lifecourse Epidemiology Unit, Southampton General Hospital, Tremona Road, Southampton SO16 6YD, UK;
ktp@mrc.soton.ac.uk

Accepted 17 January 2010
Published Online First 23 September 2010

ABSTRACT

Objectives It has been argued that in case–control studies, controls should be drawn from the base population that gives rise to the cases. In designing a study of occupational injury and risks arising from long-term illness and prescribed medication, we lacked data on subjects' occupation, without which employed cases (typically in manual occupations) would be compared with controls from the general population, including the unemployed and a higher proportion of white-collar professions. Collecting the missing data on occupation would be costly. We estimated the potential for bias if the selection rule were ignored.

Methods We obtained published estimates of the frequencies of several exposures of interest (diabetes, mental health problems, asthma, coronary heart disease) in the general population, and of the relative risks of these diseases in unemployed versus employed individuals and in manual versus non-manual occupations. From these we computed the degree of over- or underestimation of exposure frequencies and exposure ORs if controls were selected from the general population.

Results The potential bias in the OR was estimated as likely to fall between an underestimation of 14% and an overestimation of 36.7% (95th centiles). In fewer than 6% of simulations did the error exceed 30%, and in none did it reach 50%.

Conclusions For the purposes of this study, in which we were interested only in substantial increases in risk, the potential for selection bias was judged acceptable. The rule that controls should come from the same base population as cases can justifiably be broken, at least in some circumstances.

What this paper adds

- It is often said that in case–control studies, controls should be drawn from the base population that gives rise to the cases, but practical considerations sometimes dictate otherwise and a question therefore arises as to the likely magnitude of bias.
- We illustrate, for one such study, how reasonable quantitative estimates of the potential extent of bias have informed competing choices in study design.
- In our illustration, the likely potential bias was deemed acceptably small, suggesting that the selection rule can justifiably be broken, at least in some circumstances.

It has been argued that in the proper design of a case–control study, controls should be drawn from the base population that gives rise to the cases.[1] As many standard textbooks explain, the basic parameter of interest, the OR, involves a comparison of relative exposure frequency (or more strictly, exposure odds) in cases and in the population at risk of becoming cases.[2 3] The purpose of the control group is thus to give representative information on the exposure(s) of interest in the population at risk.

This is better assured (although not guaranteed) if cases and controls are ascertained from the same, discrete, well-defined study population. For example, in investigating the relationship between shift work and ischaemic heart disease (IHD), McNamee *et al*[4] focused on a cohort of workers from a particular company who started work at age ≤50 years between 1 January 1950 and 31 December 1992; the cases were cohort members who died from IHD at age ≤75 years during this period, while controls were chosen from living cohort members individually matched to cases by age and date of hire. Cases and controls were compared for their exposure to shift work as documented in company records. Because sampling was 'nested' within a well-defined occupational cohort, the precondition that controls should be liable to be identified as cases in the event of dying from IHD (ie, be at risk) was easily met.

The objective that exposure data be 'representative' requires further that the selection process for controls should be independent of the exposures of interest. In the above study of shift workers, we have no expectation that the selection algorithm would have systematically led to an erroneous estimate of shift work frequency among controls relative to all non-cases within defined matching strata.

Practical considerations, however, may sometimes mandate departures from the ideal. For example, in hospital-based case–control studies, controls are sampled from hospital patients with health problems other than the disease of interest.[5] An advantage of this method is that the recruitment of controls may be cheaper, and response rates higher. However, there is a danger that the exposures of controls might not accurately represent those in the population at risk of becoming cases. For example, if the focus of a study were risks from smoking, then, as many hospital-treated diseases are smoking-related, it can be appreciated that careless selection of controls might lead to an overestimate of exposure frequency in the at-risk population and an underestimate of the corresponding OR.

These issues in control selection have been well covered elsewhere.[1–4] In this paper we describe another instance in which convenience, costs and practical considerations conflict with the ideal. We illustrate for our example how reasonable quantitative estimates of the potential extent of bias can inform competing choices in study design.

FORMULATION OF THE PROBLEM

Increasingly, as response rates to other forms of investigation have fallen,[6] researchers have looked to exploit routinely collected datasets, some of which are amenable to case–control analysis. In the UK, one of these, the General Practice Research Database (GPRD), offers a log of all consultation episodes associated with significant events, illnesses or medical activity (diagnosis, referral, prescription, etc) among patients from some 370 participating general practices (an estimated 3 000 000 episodes of care covering 6% of all residents of England and Wales).[7] This resource with its large sample size and its wealth of routinely collected health and prescription data has been successfully exploited in numerous pharmaco-epidemiological studies of case–control design.[8] However, some variables of interest are typically missing, including occupational history.

We identified a study question of high policy relevance that we wished to address using the GPRD database. The populations of westernised countries are ageing. In future, therefore, the frequency of common age-related health conditions is likely to rise among the workforce, as is the proportion of workers taking prescribed medicines. Potentially, certain widely used medicines that impair arousal, concentration, cognition and psychomotor performance, and some common illnesses that result in sudden incapacity, impaired judgement or sensory deficit could increase the risk of accidental injury at work. But which drugs and diseases, by how much, in what circumstances, and with what consequences? The British government has announced strategic plans to maximise job retention rates among experienced older workers, but in delivering these plans employers require an evidence base to manage injury risks, the aim being to ensure safe job placement while at the same time avoiding needless restriction of job opportunities. However, when we conducted a systematic review on the topic[9] we found few relevant data, both overall and by type of injury (eg, fractured femur) and external cause (eg, fall). And we identified a need to improve upon cross-sectional studies with self-reported exposures and self-reported outcomes, by mounting investigations with objective measure of outcome and documented timing of exposures (to counter worries about common instrument reporting bias and reverse causation).[9]

The GPRD database overcame some of these limitations and fulfilled several requirements for a case–control analysis of occupational injury risk, co-morbidity and medication. It allowed an operational case definition (namely, male patients with a consultation episode for an injury coded as occupational, or involving plant or off-road vehicles or machinery or tools likely to be used only at work); and for each case, plentiful controls could be identified who were well matched by age, sex and general practice. A preliminary scoping exercise suggested that we would find some 1700 cases, to whom we could match 8500 controls. For each injury we could establish relevant exposure parameters, including the diagnostic Read code and date of first consultation; all prescriptions, with dates, within the 24 months preceding the event; and all diagnoses, with dates, preceding the event. We thus envisaged an analysis to establish the frequency and main reasons for consultation in the 24 months before injury consultation, the frequency of prescribing over this time, the main prescribed drugs and relative exposure odds of various illnesses and treatments in cases versus controls. As risks could vary according to time since first prescription of a drug or first onset of a new illness, so analysis could encompass various exposure time windows. Several aspects of confounding could be addressed through the matching algorithm (age, sex, geographical area) or via proxy measures available within the health-rich dataset (eg, alcoholic liver disease as a proxy for alcohol misuse).

Unfortunately, occupation was poorly recorded in the database, which raised concerns of the kind outlined in our introduction. Specifically, cases of occupational injury must necessarily come from the employed subfraction of the study population, whereas controls—in the absence of employment information—would be drawn from the whole population, among whom a proportion would be unemployed and not at risk of occupational injury. Also, cases would be more likely than employed controls to come from manual occupations, as the potential for occupational injury is greater in blue-collar work. Bias could arise if controls over-represented the prevalence of diseases and treatments that prevent work and are more common in the unemployed, or if they under-represented the (generally worse) health characteristics of manual workers. Finally, although practical experience suggests that such selection applies to only a few high-risk jobs, in theory people with health problems could be excluded from jobs with higher injury potential, and if these jobs were less common in controls then any risks of injury from ill health would tend to be underestimated. It should be noted that these potential biases, which relate to representativeness of exposure information among controls, do not all operate in the same direction.

The missing information could only be obtained at a cost. To contact study subjects and to ascertain their employment status by a questionnaire or interview was feasible but would carry significantly higher administrative costs and effort, a need for more elaborate ethics permissions and suitably anonymised third party mailings by collaborators with data control, and the potential for one bias (related to non-response) to be substituted for another. Some of the economic advantages of a routine publicly available dataset would be lost.

The case series method of analysis,[10] which compares the relative incidence of events of interest only among cases (in time windows of exposure and non-exposure), might seem to offer an attractive alternative. Each case would provide his or her own reference information. Since the technique is based solely on the experience of cases, this would circumvent any concern about differences in work and employment experience that arose from differences in case and referent sampling frames. However, the method is only suited to short-term exposures that impact on risk for a limited time period, such as acute intercurrent illnesses, exacerbations of pre-existing disease and newly prescribed treatments (for which purposes we intend using it). Over the much longer time frames of chronic illness and long-term treatment, potential exists for employment conditions to alter markedly within individuals. For such long-term exposures, the case–control design is still the preferred choice.

Faced with this dilemma, we decided to assess quantitatively the potential bias arising if controls were selected *without* employment information. How much would it matter that cases came from a subfraction of the population from which controls were sampled, breaking the rule on control selection often repeated in standard textbooks? We addressed this practical question focusing on four common exposures that would be of interest in our hypothetical case–control study, namely diabetes, anxiety-depression, asthma and coronary heart disease.

METHODS AND RESULTS

Considering the question, 'how great is the potential bias when controls are sampled from whole practice lists, rather than patients in work?', the logic that underlies quantitative estimation is as follows:

1. The exposure prevalence we wish to estimate using data from the controls is the prevalence that would apply in male patients from the GPRD who are in work. Let this be 'p'.
2. The prevalence in non-working controls, who represent a minority of all controls, will be higher by a multiple which is the RR of being exposed in the unemployed versus the employed.
3. The expected prevalence in our sample ('y') will be the weighted average of that in each subgroup, employed and unemployed, where the weighting factors are determined by the prevalence of unemployment.

To give an example:

- Suppose that diabetes is three times more common in the unemployed than the employed (3p rather than p);
- Suppose that the unemployment rate in men of working age is 5%. Then:
- y=0.95p+0.05(3p)=1.10p

In this example 'y' overestimates the true value of 'p' by 10%.

We obtained estimates of 'y' from a previously published analysis of the GPRD, which covered 1 007 913 men (employed and unemployed) registered with 288 practices in England and Wales during 1996.[9] Estimated RRs for exposures of interest (diabetes, mental health problems, asthma and coronary heart disease) in unemployed versus employed men were chosen following a brief literature review (details available on request) for their congruence with the published data[10–20] and to ensure that our assumptions were realistically founded.

In table 1 we have solved for 'p' to estimate the true prevalence in working controls, and present information on the extent to which 'y' would overestimate 'p' and the impact this would have on our mooted study, assuming that a given exposure truly increases the odds of a work-related injury by a factor of 2 or 3. (The method by which these figures are derived is illustrated separately in appendix 1 for one of the row items in table 1.)

We then repeated the process using RRs for exposures in manual versus non-manual workers, to illustrate separately the likely bias arising from this source (table 2).[24–37] The logic in calculation is similar but not identical:

- In the absence of unemployment, the expected prevalence of exposure in our sample ('y') would be the weighted average of that in each subgroup, manual and non-manual, where the weighting factors would be determined by the prevalence of manual work.
- Suppose that coronary heart disease is 1.24 times more common in manual than in non-manual workers (p rather than p/1.24);
- Suppose that the manual rate among employed men is 26.5%. Then:
- y=0.735p+0.265(1.24)p=1.0636p

'p'=y/1.0636. According to published estimates, y is 3.89%.[9] Hence, p=3.89/1.0636=3.6574; but the value of interest is 1.24p, *the prevalence in manual controls*, which equals 4.535%. 'y' would underestimate this by (4.35−3.89)/3.89=16.5%.

It may be seen from tables 1 and 2 that the potential for bias in the OR would generally be less than 20%. It was estimated as likely to fall between an underestimation of 14% and an overestimation of 36.7% (95th centiles). In fewer than 6% of simulations did the error exceed 30%, and in none did it reach 50%. This is comparable to, or smaller than the bias which might arise from incomplete response had such a case–control study been undertaken by attempting to contact patients directly.[38 39]

For simplicity, the analysis in table 2 assumes that all the cases come from manual occupations; in reality some occupational

Table 1 Potential for bias in estimating the OR when controls include the unemployed rather than being restricted to workers

RR*	Prevalence (%) among controls in our sample (y)[1]	Prevalence (%) in workers (solve for 'p')	% By which y overestimates p	Expected OR (vs 2.0)†	Expected OR (vs 3.0)†
Diabetes[12–14]					
2.00	1.59=(0.921×p)+(0.079×2p)	1.474%	7.9%	1.85	2.78
2.50	1.59=(0.921×p)+(0.079×2.5p)	1.422%	11.9%	1.79	2.68
3.00	1.59=(0.921×p)+(0.079×3p)	1.373%	15.8%	1.72	2.58
Mental health[15–18]					
2.00	3.83=(0.921×p)+(0.079×2p)	3.550%	7.9%	1.85	2.77
2.50	3.83=(0.921×p)+(0.079×2.5p)	3.424%	11.9%	1.78	2.67
3.00	3.83=(0.921×p)+(0.079×3p)	3.307%	15.8%	1.72	2.58
Prescribed antidepressants[16]					
3.00	3.47=(0.921×p)+(0.079×3p)	2.997%	15.8%	1.72	2.58
Prescribed anxiolytics[16]					
2.50	3.03=(0.921×p)+(0.079×2.5p)	2.709%	11.9%	1.78	2.67
Asthma[19 20]					
1.20	6.55=(0.921×p)+(0.079×1.2p)	6.448%	1.6%	1.97	2.95
1.50	6.55=(0.921×p)+(0.079×1.5p)	6.301%	4.0%	1.92	2.88
2.00	6.55=(0.921×p)+(0.079×2p)	6.070%	7.9%	1.84	2.77
Coronary heart disease[21 22]					
1.50	3.89=(0.921×p)+(0.079×1.5p)	3.742%	4.0%	1.92	2.88
2.00	3.89=(0.921×p)+(0.079×2p)	3.605%	7.9%	1.85	2.77
3.00	3.89=(0.921p)+(0.079×3p)	3.359%	15.8%	1.72	2.58

The table assumes an unemployment rate of 7.9%, which is the average for men aged 16–64 years during 1987–2007 in Britain.[23]
*RR of the exposure in question (eg, diabetes) in the employed versus the unemployed.
†Expected OR of occupational injury in those with the exposure versus those without, assuming true ORs of 2.0 and 3.0, respectively. The method by which these figures are calculated is illustrated in appendix 1.

Table 2 Potential for bias in estimating the OR when controls include non-manual workers and cases are all manually employed

RR*	Definition of 'manual' occupation	Prevalence (%)[1] among controls in our sample (y)	Prevalence (%) in manual workers (solve for 'p', then multiply by RR)	% By which y underestimates	Expected† OR (vs 2.0)	Expected† OR (vs 3.0)
Coronary heart disease						
1.24	Manual (IV/V, IIIM) vs non (I/II, IIIN) (MI, ischaemic ECG)[24]	3.89=(0.735×p)+(0.265×1.24p)	4.535%	16.6%	2.35	3.52
1.40	Manual (IV/V, IIIM) vs non (I/II, IIIN) (recall of IHD)[24]	3.89=(0.735×p)+(0.265×1.40p)	4.924%	26.6%	2.56	3.84
1.31	Manual (IV/V, IIIM) vs non (I/II, IIIN)[25]	3.89=(0.512×p)+(0.488×1.31p)	4.428%	13.8%	2.29	3.43
1.30 to 1.33	Manual versus professional[26]	3.89=(0.65×p)+(0.35×1.30p) 3.89=(0.65×p)+(0.35×1.33p)	4.576% to 4.638%	17.6% to 19.2%	2.37 to 2.40	3.55 to 3.60
1.52	Lower supervisory, technical, routine versus managerial, professional, intermediate, own account[27]	3.89=(0.634×p)+(0.366×1.52p)	4.967%	27.7%	2.58	3.87
Asthma						
1.08	Manual (IV/V, IIIM) vs non (I/II, IIINM, missing) (20–44 year olds)[28]	6.55=(0.637×p)+(0.363×1.08p)	6.880%	5.0%	2.11	3.16
1.05 to 1.20	Low SES versus high SES[29]	6.55=(0.675×p)+(0.325×1.05p) 6.55=(0.675×p)+(0.325×1.20p)	6.768% to 7.380%	3.6% to 13.7%	2.07 to 2.27	3.11 to 3.41
1.01	Lower supervisory, technical, routine versus managerial, professional, intermediate, own account (doctor-diagnosed asthma)[30]	6.55=(0.557×p)+(0.443×1.01p)	6.597%	0.7%	2.02	3.02
1.38	Lower supervisory, technical, routine versus managerial, professional, intermediate, own account (wheeze, past 12 months)[30]	6.55=(0.557×p)+(0.443×1.38p)	7.749%	18.3%	2.40	3.60
Diabetes						
1.07	Low versus middle or high present occupational position (impaired glucose tolerance)[32]	1.59=(0.655×p)+(0.345×1.07p)	1.661%	4.5%	2.09	3.14
1.97	Low versus middle or high present occupational position (type 2 diabetes)[32]	1.59=(0.655×p)+(0.345×1.97p)	2.347%	47.6%	2.97	4.46
1.24	Lower supervisory, technical, routine versus managerial, professional, intermediate, own account[33]	1.59=(0.604×p)+(0.396×1.24p)	1.800%	13.3%	2.27	3.41
1.30	Manual (IV/V, III) vs non (I/II)[34]	1.59=(0.63×p)+(0.37×1.30p)	1.864%	17.2%	2.35	3.53
Mental health						
1.04	Manual (IV/V, IIIM) vs non (I/II, IIIN) (neurotic disorder)[35]	3.83=(0.525×p)+(0.475×1.03p)	3.899%	1.8%	2.04	3.06
1.59	Non-skilled versus skilled (depression)[36] ‡	3.83=(0.698×p)+(0.302×1.59p)	5.167%	34.9%	2.74	4.10
0.95	No access to car/van versus access	3.83=(0.601×p)+(0.3992×0.95p)	3.713%	−3.1%	1.94	2.90
1.19	Rented accommodation versus not	3.83=(0.72×p)+(0.28×1.20p)	4.327%	13.0%	2.27	3.41
1.29	Not saving from income versus saving	3.83=(0.421×p)+(0.579×1.30p)	4.230%	10.5%	2.22	3.33
1.40	House repairs versus not (GHQ minor psychiatric ill-health)[37] ‡	3.83=(0.746×p)+(0.254×1.40p)	4.867%	27.1%	2.57	3.85

The definition of 'manual' occupation varied between publications, and was seldom dichotomous. To simplify, we regroup the data from source reports according to the definition in the second column. Where appropriate we derive a RR (first column) and prevalence of manual occupation (used in third column) based on the new groupings from the published data (calculations available on request).

*RR of the exposure in question (eg, diabetes) in lower versus higher social class.

†Expected OR of occupational injury in those with the exposure versus those without, assuming true ORs of 2.0 and 3.0, respectively. The method by which these figures are calculated is illustrated in appendix 1.

‡Data for men were not separately disaggregated; in Weich and Lewis,[37] however, RRs were adjusted for sex.

GHQ, General Health Questionnaire; IHD, ischaemic heart disease; MI, myocardial infarction; SES, socio-economic status.

injuries would arise in non-manual workers. Thus, table 2 somewhat overstates the likely bias.

It should be noted that the direction of bias is different in the two tables, leading to an underestimate in table 1 and an overestimate in table 2. In practice, controls would represent a mixture of manual workers, non-manual workers and the unemployed. Hence, the actual bias would be expected to lie between the values in the two tables.

DISCUSSION

In designing this case–control protocol, practical considerations encouraged us to violate a well-rehearsed axiom of control selection. A trade-off then existed between cost and possible loss of internal validity. Although different input assumptions would yield different values, with the quantitative assumptions presented we judged the potential bias as acceptable, particularly when set against the alternative bias that might arise from attempted patient contact and incomplete response.

The method by which we estimated the potential extent of bias in this planning exercise was somewhat similar to that which has been applied when estimating possible impacts of uncontrolled confounding after data collection has been completed.[40 41] This reflects a conceptual overlap between selection bias and confounding. Thus, the controls in our proposed study would include unemployed people, whose health status and use of medication is likely to differ

systematically from that of those in employment. Viewed one way, the resultant unrepresentativeness of controls could be classed as a selection bias. Alternatively, however, employment status could be considered as a confounding variable, associated with the risk factors of interest (health status and use of medication), and independently determining the risk of occupational injury.

The findings of our analysis are not wholly unexpected, since the potential for bias must reflect the weighted average of risks of exposure in component subgroups (employed versus unemployed on the one hand and manual versus non-manual on the other). In relation to unemployment, it would be limited because unemployed subjects are in the minority, and in the case of type of work by the moderate RRs between manual and non-manual occupations (although RRs are greater at the extremes (social class V versus I), these represent only a fraction of the whole population).

When a study is being planned, the potential extent of bias that is deemed tolerable will depend on the use that will be made of its findings. This is analogous to consideration of the scope for random error—the statistical power that is required of a study will vary according to the context. In the example that we have given, our interest was only in detecting substantially increased risks of occupational injury. Minor hazards would have no impact on employment decisions. For this reason, a possible error of 20% was judged acceptable. However, in other circumstances, the investigator might be interested in discriminating much smaller deviations from the null, in which case a possible error of 20% in risk estimates might be considered unacceptable.

We conclude that in this particular study the absence of data on employment status, although a drawback, would not be a critical limitation. Studies involving other exposures and outcomes would need to be considered on their individual merits. However, simple calculations and externally published data can be used to obtain an estimate of the potential for bias. The selection rule can justifiably be broken, at least in some circumstances.

Acknowledgements We would like to thank Dr Clare Harris for her help in compiling some of the published risk estimates on which calculations are based, and for her careful proof-reading of the manuscript.

Competing interests None.

Funding The authors are in receipt of core research funding from the Medical Research Council, UK.

Contributors KP and DC conceived the idea; KP wrote the first draft of the paper and DC assisted in revision. KP and MK were responsible for the calculations. KP acts as guarantor for the work. All the authors have read and approved the manuscript.

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES

1. **Wacholder S,** McLaughlin JK, Silverman DT, *et al*. Selection of controls in case-control studies: I Principles. *Am J Epidemiol* 1992;**135**:1019—28.
2. **Schlesselman JJ.** *Case-control Studies: Design, Conduct, Analysis*. Oxford University press: Oxford, 1982.
3. **Rothman KJ.** *Modern Epidemiology*. Boston: Little, Brown and Company, 1986.
4. **McNamee R,** Binks K, Jones S, *et al*. Shift work and mortality from ischaemic heart disease. *Occup Environ Med* 1996;**53**:367—73.
5. **Wacholder S,** Silverman DT, McLaughlin JK, *et al*. Selection of controls in case-control studies: II. Types of Controls. *Am J Epidemiol* 1992;**135**:1029—41.
6. **De Heer WF,** Israëls AZ, eds. Response trends in Europe. *American Statistical Association; Proceedings of the Section on Survey Research Methods* 1992:92—101.
7. **Wally T,** Mantgani A. The UK general practice research database. *Lancet* 1997;**350**:1097—9.
8. **Medicines & Healthcare products Regulatory Agency.** GPRD bibliography. http://www.gprd.com/bibliography/ (accessed 1 Jul 2009).
9. **Palmer KT,** Harris EC, Coggon D. Chronic health problems and risk of accidental injury in the workplace: a systematic literature review. *Occup Environ Med* 2008;**65**:757—64.
10. **Farrington CP,** Nash J, Miller E. Case series analysis of adverse reactions to vaccines: a comparative evaluation. *Am J Epidemiol* 1996;**143**:1165—73.
11. **Office of National Statistics.** *Key Health Statistics from General Practice*. Series MB6 no. 1. London: ONS, 1996. ISBN I 85774-2737.
12. **Kraut A,** Walld R, Tate R, *et al*. Impact of diabetes on employment and income in Manitoba, Canada. *Diabetes Care* 2001;**24**:64—8.
13. **Robinson N,** Yateman NA, Protopapa LE, *et al*. Unemployment and diabetes. *Diabet Med* 1989;**6**:797—803.
14. **Robinson N,** Yateman NA, Protopapa LE, *et al*. Employment problems and diabetes. *Diabet Med* 1990;**7**:16—22.
15. **Comino EJ,** Harris E, Silove D, *et al*. Prevalence, detection and management of anxiety and depressive symptoms in unemployed patients attending general practitioners. *Aust N Z J Psychiatry* 2000;**34**:107—13.
16. **Comino EJ,** Harris E, Chey T, *et al*. Relationship between mental health disorders and unemployment status in Australian adults. *Aust N Z J Psychiatry* 2003;**37**:230—5.
17. **Khlat M,** Sermet C, Le Pape A. Increased prevalence of depression, smoking, heavy drinking and use of psycho-active drugs among unemployed men in France. *Eur J Epidemiol* 2004;**19**:445—51.
18. **Wittchen HU,** Zhao S, Kessler RC, *et al*. DSM-III-R generalized anxiety disorder in the National Comorbidity Survey. *Arch Gen Psychiatry* 1994;**51**:355—64.
19. **Sibbald B,** Anderson HR, McGuigan S. Asthma and employment in young adults. *Thorax* 1992;**47**:19—24.
20. **Kogevinas M,** Anto JM, Tobias A, *et al*. Respiratory symptoms, lung function and use of health services among unemployed young adults in Spain. Spanish Group of the European Community Respiratory Health Survey. *Eur Respir J* 1998;**11**:1363—8.
21. **Yarnell J,** Yu S, McCrum E, *et al*. PRIME study group. education, socioeconomic and lifestyle factors, and risk of coronary heart disease: the PRIME Study. *Int J Epidemiol* 2005;**34**:268—75.
22. **Cook DG,** Cummins RO, Bartley MJ, *et al*. Health of unemployed middle-aged men in Great Britain. *Lancet* 1982;**1**:1290—4.
23. **National Statistics Online.** Labour market statistics, time series data - unemployment by age and duration, http://www.statistics.gov.uk/statbase/tsdtables1.asp?vlnk=lms (accessed 1 Jul 2009).
24. **Pocock SJ,** Shaper AG, Cook DG, *et al*. Social class differences in ischaemic heart disease in British men. *Lancet* 1987;**2**:197—201.
25. **Department of Health.** Health Survey for England 1998: cardiovascular disease. *Age standardised prevalence in men aged 35+ years*. The Stationery Office, 1999. http://www.archive.official-documents.co.uk/document/doh/survey98/hse-02.htm#2.4 (accessed 1 Jul 2009).
26. **Bennett S.** Socioeconomic inequalities in coronary heart disease and stroke mortality among Australian men, 1979—1993. *Int J Epidemiol* 1996;**25**:266—75.
27. **Emberson JR,** Whincup PH, Morris RW, *et al*. Social class differences in coronary heart disease in middle-aged British men: implications for prevention. *Int J Epidemiol* 2004;**33**:289—96.
28. **National Centre for Social Research, Department of Epidemiology and Public Health.** *Health Survey for England 2003. Vol. 1 Cardiovascular Disease*. The Stationery Office, 2004. http://www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsStatistics/DH_4098712 (accessed 1 Jul 2009).
29. **Basagaña X,** Sunyer J, Kogevinas M, *et al*; European Community Respiratory Health Survey. Socioeconomic status and asthma prevalence in young adults: the European Community Respiratory Health Survey. *Am J Epidemiol* 2004;**160**:178—88.
30. **Bråbäck L,** Hjern A, Rasmussen F. Social class in asthma and allergic rhinitis: a national cohort study over three decades. *Eur Respir J* 2005;**26**:1064—8.
31. **Department of Health.** *Health Survey for England 2001: Respiratory Symptoms, Atopic Conditions and Lung Function*. The Stationery Office, 2001. http://www.archive2.official-documents.co.uk/document/deps/doh/survey01/rsac/rsac.htm (accessed 1 Jul 2009).
32. **Agardh EE,** Ahlbom A, Andersson T, *et al*. Socio-economic position at three points in life in association with type 2 diabetes and impaired glucose tolerance in middle-aged Swedish men and women. *Int J Epidemiol* 2007;**36**:84—92.
33. **National Centre for Social Research, Department of Epidemiology and Public Health.** *Health Survey for England 2003. Vol. 2 Risk Factors for Cardiovascular disease*. The Stationery Office, 2004. http://www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsStatistics/DH_4098712 (accessed 1 Jul 2009).
34. **Larranaga I,** Arteagoitia JM, Rodriguez JL, *et al*; Sentinel Practice Network of the Basque Country. Socio-economic inequalities in the prevalence of Type 2 diabetes, cardiovascular risk factors and chronic diabetic complications in the Basque Country, Spain. *Diabet Med* 2005;**22**:1047—53.
35. **Lewis G,** Bebbington P, Brugha T, *et al*. Socio-economic status, standard of living, and neurotic disorder. *Lancet* 1998;**352**:605—9.
36. **Cheng TA.** A community study of minor psychiatric morbidity in Taiwan. *Psychol Med* 1988;**18**:953—68.
37. **Weich S,** Lewis G. Material standard of living, social class, and the prevalence of the common mental disorders in Great Britain. *J Epidemiol Community Health* 1998;**52**:8—14.

38. **Lahkola A,** Salminen T, Auvinen A. Selection bias due to differential participation in a case–control study of mobile phone use and brain tumors. *Ann Epidemiol* 2005;**15**:321–5.
39. **Madigan MP,** Troisi R, Portischman N, *et al*. Characteristics of respondents and non-respondents from a case-control study of breast cancer in younger women. *Int J Epidemiol* 2000;**29**:793–8.
40. **Greenland S.** Basic methods for sensitivity analysis and external adjustment. In: Rothman KJ, Greenland S, eds. *Modern epidemiology*. 2nd edn. Philadelphia: Lippincott-Raven, 1998;Chapter 19:343–7.
41. **Arah OA,** Chiba Y, Greenland S. Bias formulas for external adjustment and sensitivity analysis of unmeasured confounders. *Ann Epidemiol* 2008;**18**:637–46.

APPENDIX 1

The potential for bias in estimation of ORs: a worked example

Consider the example of diabetes and the effect of unemployment status, with the following input assumptions….

- The true OR we seek to estimate (odds of occupational injury in those with diabetes versus those without)=2.0
- The RR of diabetes in employed versus unemployed men=3.0
- The estimate of prevalence of diabetes in our controls (y)=1.59%[9]
- We planned to study 1700 cases and 8500 controls….

RR	Prevalence (%) among controls in our sample (y)	Prevalence (%) in workers (solve for 'p')	Expected OR (vs 2.0)
Diabetes			
3.00	1.59=(0.921×p)+(0.079×3p)	1.373%	1.72

The extract from table 1 (above) shows that the estimated prevalence of diabetes in working controls (p) is 1.373%, and that the OR of 2.0 can be expected to be biased downwards to 1.72. This last figure is derived as follows:

If all the controls were workers, 1.373% of 8500 that is 116.705 (without rounding) would be diabetics and the remainder (8383.295) would not.

In fact, as our controls include some unemployed men, and as a whole have a prevalence of 1.59%, we estimate in error that 135.15 controls would have diabetes and 8364.85 would not.

Imagine first the 'true' 2×2 table, confined to workers, among whom the true OR for injury is 2.

Worker controls			
	Diabetes?		
Injury?	Yes	No	All
Yes	A	(1700−A)	1700
No	116.705	8383.295	8500

This table has one unknown, but OR=2. Thus, (8383.295×A)/(116.705×(1700−A))=2.

Solving for 'A' gives a value of 46.05:

Worker controls			
	Diabetes?		
Injury?	Yes	No	All
Yes	46.05	1653.95	1700
No	116.705	8383.295	8500

Using 'all' controls rather than 'worker' controls will alter the bottom row of this table as follows:

All controls			
	Diabetes?		
Injury?	Yes	No	All
Yes	46.05	1653.95	1700
No	135.15	8364.85	8500

Thus, instead of an OR of 2, the estimated OR would become: (46.05×8364.85)/(135.15×1653.95)=1.723.

Corrections

NO2 and children' s respiratory symptoms in the PATY study. **Pattenden S,** Hoek G, Braun-Fahrlander C *et al Occup Environ Med* 2006;**63**:828–835. This article was published with an incorrect doi of 10.1136/oem.2006.025213. The correct doi is 10.1136/oem.2005.025213.

Occup Environ Med 2010:**67**:877. doi:10.1136/oem.2005.025213

Valentini E, Ferrara M, Prasaghi F *et al*. Systematic review and meta-analysis of psychomotor effects of mobile phone electromagnetic fields. *Occup Environ Med* 2010;**67**:708–716. The citation in this review contains an error. The fourth author is De Gennaro L, not Gennaro LD.

Occup Environ Med 2010:**67**:877. doi:10.1136/oem.2009.047027corr1