

Bias in psychiatric case–control studies

Literature survey

WILLIAM LEE, JONATHAN BINDMAN, TAMSIN FORD, NICK GLOZIER, PAUL MORAN, ROBERT STEWART and MATTHEW HOTOPF

Background: Case–control studies are vulnerable to selection and information biases which may generate misleading findings.

Aims To assess the quality of methodological reporting of case–control studies published in general psychiatric journals.

Method All the case–control studies published over a 2-year period in the six general psychiatric journals with impact factors of more than 3 were assessed by a group of psychiatrists with training in epidemiology using a structured assessment devised for the purpose. The measured study quality was compared across type of exposure and journal.

Results The reporting of methods in the 408 identified papers was generally poor, with basic information about recruitment of participants often absent. Reduction of selection bias was described best in the ‘pencil and paper’ studies and worst in the genetic studies. Neuroimaging studies reported the most safeguards against information bias. Measurement of exposure was reported least well in studies determining the exposure with a biological test.

Conclusions Poor reporting of recruitment strategies threatens the validity of reported results and reduces the generalisability of studies.

Declaration of interest None.

Many studies in psychiatry compare biological, psychological, or social variables between healthy controls and individuals with psychiatric disorder. These studies are conceptually similar to the case–control study of epidemiology in that the participants are selected according to the presence or absence of a disorder. The two main sources of bias in case–control studies are selection bias and information bias. Selection bias exists where exposure status has a non-random effect on the selection of either cases or controls. The choice of the control group is crucial in this respect, since it functions to represent the level of exposure within the general population from which the cases have been identified. Information bias includes recall bias (where the participants’ illness experience systematically affects recall) and observer bias (where the knowledge the investigator has about the study hypothesis and of participants’ case or control status influences the assessment of the parameter under study). Case–control studies are an important source of evidence for many areas of mental health research. In a survey of papers published in leading non-specialist psychiatric journals, we evaluated the reported quality of the methods of case–control studies in psychiatry and evaluated the extent to which measures were taken to avoid these potential biases.

METHOD

Identification of studies

We hand-searched general psychiatric journals with an impact factor greater than 3 in 2001 from January 2001 to December 2002 inclusive. Studies were included if they compared participants with psychiatric disorders with healthy controls on any variable. Post-mortem studies were excluded, as were twin, co-twin and affected sibling designs.

Assessment of studies

We devised a data extraction form to describe the general characteristics of the paper, the selection of cases and controls, and the methods used to reduce information bias. We recorded the parameter compared between groups, the type and number of cases and the type and number of controls. If more than two diagnoses were studied we assigned the most numerous group to the cases, and did not collect details of other diagnostic groups. We also recorded details of individual matching and, if matching was performed, whether a matched analysis was used.

To examine selection bias we recorded details of the clinical setting where recruitment took place and whether the denominator from which cases were selected was described. For example, studies that reported recruiting patients with a specific diagnosis from consecutive series of new referrals to a service, and gave details of the total number of patients eligible, would score for both items. We collected information on whether new (incident) cases were used, descriptions of the duration of illness, and the use of medication for disorders in which these data are relevant. We focused on the process by which recruitment was undertaken – in particular whether information was supplied on the total number of potential participants who were approached, the numbers of participants and non-participants, and whether differences between participants and non-participants were described. We also assessed whether inclusion and exclusion criteria were described in sufficient detail for the study to be replicated by other researchers. We recorded whether controls were recruited from students or employees of the organisation where the research was performed; whether they were selected from a defined population; whether they were recruited from advertisements; how many were approached; whether the differences between participant and non-participant controls were described; and whether similar exclusion criteria were applied to both cases and controls.

To assess information bias, we recorded whether the determination of exposure status had been carried out in a comparable way for both cases and controls and whether the investigators performing ratings had been masked to the participants’ illness status.

We piloted the rating scale by testing the interrater reliability of each item for

22 papers: The raters (J.B., T.F., N.G., M.H., P.M. and R.S.) are members of the Royal College of Psychiatrists and all have postgraduate qualifications in epidemiology. All papers published in January 2001 (or the next chronological paper if no paper was identified from that month) were rated by all six raters. The answers were compared formally and a consensus reached at a meeting on items where differences were identified, resulting in a rater manual. Each rater then used this scheme to rate a further 47–64 papers.

We categorised the papers into four broad groups, depending on the techniques used to acquire the 'exposure' data:

- (a) neuroimaging: structural or functional imaging;
- (b) biological: properties of samples taken from the participants (e.g. blood, saliva) or biometrics;
- (c) pencil and paper: psychometric tests or questionnaires, either self-completed or interviewer-rated;
- (d) genetic.

To allow for comparison of the overall measured quality of the papers, we created three simple scales in which the scores consisted of the number of questionnaire items with answers indicative of good practice for the nine items concerning selection bias of cases, the six items concerning selection bias of controls, and the two items concerning information bias. We compared the measured quality of the papers using these scales in relation to research topic and the journal of publication.

RESULTS

Interrater reliability

Twenty-two (5%) of the 408 papers were rated by all six of the raters. Seven of the papers were neuroimaging papers, eight were biological, six were pencil and paper, and one was a genetics paper. Of the 17 questions answered by the raters, three had a kappa value of greater than 0.8, five had kappa values between 0.6 and 0.8, two had kappa values between 0.4 and 0.6 and seven had kappa values of less than 0.4. (All but one of the questions had a percentage agreement in excess of 70% and many of those with the lowest kappa values had the highest percentage agreements. Even highly reliable measures show low kappa values when the expected frequency is low, as in this case; Altman, 2003). For each item on the questionnaire, a consensus answer was reached at a meeting of the raters. A manual was devised such that the raters using the manual gave the consensus answer on retesting.

Sample

The six journals that met the inclusion and exclusion criteria are listed in Table 1. From these journals 408 papers were identified. Eligible studies represent between 2% (*Journal of Clinical Psychiatry*) and 55% (*Archives of General Psychiatry*) of all published research. Papers reporting neuroimaging studies accounted for the largest number of papers in four of the six journals, with papers involving paper and pencil tests being the most frequent in the remaining two journals (*Psychological Medicine* and *Journal of Clinical Psychiatry*). Genetic papers were the least

numerous in the sample (Table 1). Table 2 shows the study sample sizes by research area and journal. In general sample sizes were small, with a median group size of 23.5 (interquartile range 15.0–43.5). The groups were particularly small in biological and neuroimaging studies.

Selection bias

The questionnaire items concerning the clinical setting from which participants were recruited and medication use were described the most adequately, with 61% and 68% of papers respectively providing satisfactory information. Approximately half of the papers performed satisfactorily on the items concerning the use of similar exclusion criteria for cases and controls (57%) and the description of inclusion and exclusion criteria (50%). However, the reporting was particularly poor in four of the items: few of the papers fully described participants and non-participating potential cases (5%), or the differences between them (2%); similarly, information on the number of potential controls approached was rarely provided (5%), and only 1% of papers described the differences between participating controls and those who were approached to be controls but declined (Table 3). Two items (the use of students or employees of the research institution and the use of advertising for recruitment) were very frequently rated as 'unclear', indicating insufficient information was available to make a judgement. However, at least a third of all studies used advertisements to recruit controls, and at least 15% used staff or students from the research institution as controls.

Table 1 Distribution of the included case-control studies between journals and areas of research

Journal	Research area, n (%) ¹				Case-control studies n (%) ²	Total published studies n (%) ³
	Neuroimaging	Biological	Pencil and paper	Genetics		
<i>American Journal of Psychiatry</i>	66 (47)	37 (26)	34 (24)	3 (2)	140 (34)	519 (27)
<i>Archives of General Psychiatry</i>	26 (46)	18 (32)	9 (16)	4 (7)	57 (14)	104 (55)
<i>Biological Psychiatry</i>	57 (47)	43 (35)	13 (11)	9 (7)	122 (30)	469 (26)
<i>British Journal of Psychiatry</i>	10 (42)	5 (21)	9 (38)	0 (0)	24 (6)	600 (4)
<i>Journal of Clinical Psychiatry</i>	0 (0)	1 (20)	4 (80)	0 (0)	5 (1)	250 (2)
<i>Psychological Medicine</i>	10 (17)	11 (18)	39 (65)	0 (0)	60 (15)	300 (20)
Total	169 (41)	115 (28)	108 (26)	16 (4)	408 (100)	2267 (18)

1. Values in parentheses are percentages of row totals.

2. Values in parentheses are percentages of column total.

3. Values in parentheses are the proportion of papers published in each journal that are case-control studies.

Table 2 Median size and interquartile range of the largest case group in each study

	Studies, n (%)	Sample size, n Median (IQR)
Research area		
Neuroimaging	169 (41)	18 (12–30)
Biological	115 (28)	21 (15–38)
Pencil and paper	108 (26)	38.5 (23–88.5)
Genetic	16 (4)	108 (36–177.5)
Journal		
<i>American Journal of Psychiatry</i>	140 (34)	22.5 (15–41)
<i>Archives of General Psychiatry</i>	57 (14)	22 (15–40)
<i>Biological Psychiatry</i>	122 (30)	21 (14–42)
<i>British Journal of Psychiatry</i>	24 (6)	25.5 (17–88)
<i>Journal of Clinical Psychiatry</i>	5 (1)	34 (20–40)
<i>Psychological Medicine</i>	60 (15)	26 (19.5–49)
Total	408 (100)	23.5 (15–43.5)

IQR, interquartile range.

Information bias

Most (93%) papers reported that they assessed exposure status in a sufficiently

similar way for cases and controls (Table 3), but only 25% indicated that the investigators were ‘masked’ to the illness status of the participants, and in 70% of

the papers it was impossible to determine whether the investigators were ‘masked’ or not.

Matching and analysis

In 121 of the 408 studies (30%) participants were individually matched. There was no difference, either by area of research or journal, in the proportion of studies that carried out individual matching of participants. Only 30% of the studies that used this technique carried out a matched analysis. There was no significant difference in this proportion between research areas or journal of publication (not shown).

Overall quality of the papers

Studies that used pencil and paper tests showed significantly more desirable methodological features in the selection of both cases and controls than the studies in other research areas. Genetic studies were rated poorest in the selection of cases. Neuroimaging studies showed most desirable features

Table 3 Answers to items in the questionnaire used to evaluate the methodological quality of the case–control studies.

Question	Answer		
	Yes n (%) ¹	No n (%)	Unclear n (%)
Selection bias			
Cases			
Was the clinical setting used for recruitment made clear?	248 (61)	160 (39)	
Was the denominator from which cases were recruited described?	96 (24)	301 (74)	11 (3)
Were incident cases used?	44 (11)	344 (84)	20 (5)
Was duration of illness adequately described?	174 (43)	212 (52)	22 (5)
Was medication use adequately described?	277 (68)	86 (21)	45 (11)
Was adequate information given on the total number of patients approached?	43 (11)	357 (88)	8 (2)
Was information given on participants and non-participants?	20 (5)	379 (93)	9 (2)
Was information given on the differences between participants and refusers?	9 (2)	390 (96)	9 (2)
Were the inclusion and exclusion criteria described well enough to be replicable?	203 (50)	205 (50)	
Controls			
Did the study use controls who were students/employees of the research institution? ²	56 (14)	125 (31)	227 (56)
Were controls selected from an explicit sampling frame?	67 (16)	332 (81)	9 (2)
Did the study recruit through advertisements? ²	143 (35)	46 (11)	219 (54)
Were similar exclusion criteria applied for controls as for cases?	231 (57)	32 (8)	145 (36)
Was information given on number of controls approached?	21 (5)	375 (92)	12 (3)
Was adequate information given on differences between controls refusing and agreeing?	3 (1)	395 (97)	10 (2)
Information bias			
Was ‘exposure’ status performed in a sufficiently similar way?	381 (93)	8 (2)	19 (5)
Were the investigators who rated the exposure masked to participants’ status? ³	95 (25)	16 (4)	265 (70)

1. Values in parentheses are row percentages; the denominator in each row is the 408 papers included in the study.

2. Denotes that ‘no’ is the answer indicative of good methodological practice.

3. Thirty-two studies were removed from the denominator in this item because no human decision was required to rate the exposure.

Table 4 Numbers of questions in each section of the questionnaire that were answered indicating good practice, by research area and source journal

	Selection bias (cases) (0–9) Median (IQR)	Selection bias (controls) (0–6) Median (IQR)	Information bias (0–2) Median (IQR)
Research area			
Neuroimaging	3 (2–4)	1 (0–1)	1 (1–2) [†]
Biological	3 (2–4)	1 (0–2)	1 (1–1)
Pencil and paper	3 (2–5) [†]	1 (0–2.5) [†]	1 (1–1)
Genetic	2 (1–3) [†]	1 (0–1)	1 (1–1)
Journal			
<i>American Journal of Psychiatry</i>	3 (2–4)	1 (0–1)	1 (1–1)
<i>Archives of General Psychiatry</i>	3 (2–4)	1 (1–2) [†]	1 (1–2)
<i>Biological Psychiatry</i>	3 (1–4) [†]	1 (0–1) [†]	1 (1–1)
<i>British Journal of Psychiatry</i>	3.5 (1.5–5)	1 (1–2.5)	1 (1–2)
<i>Journal of Clinical Psychiatry</i>	4 (3–5)	2 (0–2)	1 (1–1)
<i>Psychological Medicine</i>	3 (2–4)	1 (0.5–2)	1 (1–1)
Total	3 (2–4)	1 (0–2)	1 (1–1)

IQR, interquartile range.

†. The median of the group is statistically significantly different from the median of the other groups making up the entire sample, using the Kruskal–Wallis test with a *P* value required for significance corrected for multiple comparisons to 0.0125 for the research areas and 0.00833 for the journals.

in the elimination of information bias (Table 4).

Papers published in *Biological Psychiatry* were rated as showing fewest desirable features in the recruitment of cases and controls. Papers published in *Archives of General Psychiatry* showed significantly superior methodology in reducing selection bias of controls compared with papers published in other journals (Table 4).

The data from our three quality rating scales are shown in histogram form in Figs 1–3.

DISCUSSION

The case–control study design is common across many areas of psychiatric research, as it is a cost-effective study design,

especially for relatively rare psychiatric outcomes such as psychotic illness. In this review, we found that the general level of methodological description was poor and many of the papers failed to include sufficient information to allow a judgement to be made about the impact of selection or information biases on the findings of the studies. Genetic studies achieved the poorest ratings in reducing selection bias, whereas pencil and paper studies achieved the best. Neuroimaging studies gave the most complete information relevant to information bias. There were few differences between journals in the reporting of measures to reduce information biases.

The recruitment of participants was not described well in most of the studies examined. This means that the generalisability of the

findings arising from these studies cannot be assessed, and that accurate replication of the study in a different population or time period becomes impossible. In case–control studies the control group functions to represent the level of exposure within the general population from which the cases have been identified, and researchers should ensure that the selection of cases and controls takes place within a defined population in as transparent and reproducible a manner as possible (Wacholder, 1995). The practice of advertising within a research institution to recruit controls – who are frequently students or staff members of that organisation – is widespread and is likely to introduce biases which may be difficult to quantify. It is not improbable that the often subtle experimental

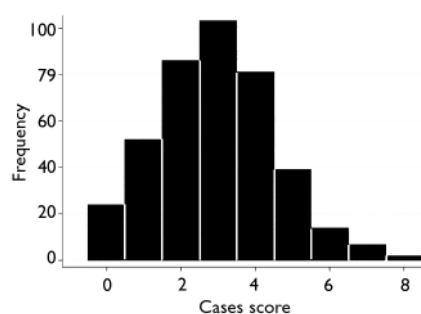


Fig. 1 Data from the nine-point rating scale assessing the quality of the recruitment of cases.

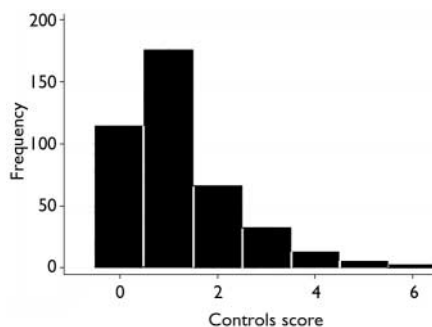


Fig. 2 Data from the six-point rating scale assessing the quality of the recruitment of controls.

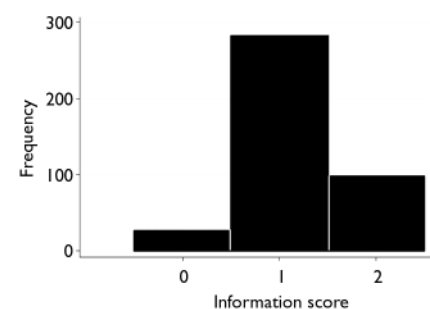


Fig. 3 Data from the two-point rating scale assessing the minimisation of information biases.

conditions devised in functional brain imaging studies may be influenced by educational level or motivation to participate in research. Further, the poor quality of reporting of the selection of cases suggests that many studies use what are effectively 'convenience' samples, which will tend to comprise the most severe and treatment-resistant cases in a service. These two opposing factors – 'super-healthy' controls and unrepresentatively ill cases – are likely to lead to an overestimate of effect sizes (Lewis & Pelosi, 1990).

The masking of raters was generally poorly reported. There are, no doubt, situations in which a parameter can be estimated without any risk of observer bias and therefore with no theoretical need for masking. However, it is difficult to determine when these situations are present. Many apparently 'hard' outcomes – such as volume of brain structures or concentrations of immune parameters – involve a good deal of measurement performed by humans and are therefore open to observer bias (Sackett, 1979). It is hard to envisage a situation where masking of those performing such ratings is not feasible, and we can think of no situation where to attempt masking would be harmful. We therefore suggest that authors have a duty either to report that masking took place or the reasons why this was unnecessary. In the majority of papers we assessed, this information was not available. Those reading the papers without a detailed knowledge of the techniques used have no idea whether observer bias is a possible explanation of the reported findings.

Unlike chance and confounding, bias cannot be readily quantified, may not be detectable and cannot be taken into account in data analysis. This means that the only opportunity to reduce the influence of bias on the results of a study is at the design phase. Problems with the methodology and reporting of randomised controlled trials were observed in the 1990s (Schulz, 1995a,b,c,1996; Hotopf *et al*, 1997; Ogunidipe *et al*, 1999). An outcome of this was the Consolidated Standards of Reporting Trials (CONSORT) statement, in which authors are required to describe their methodology according to a 22-item checklist (Altman *et al*, 2001). This has unified clinicians, academics, policy makers and the pharmaceutical industry, and is now a mandatory part of submissions of randomised controlled trials to major journals.

A number of reviews have documented many areas of scientific research where the findings of case-control studies have not been replicated in methodologically superior prospective cohort studies (Mayes *et al*, 1988; Pocock *et al*, 2004; von Elm & Egger, 2004). In psychiatry, the emerging finding that large, population-based case-control neuroimaging studies in psychosis (Dazzan *et al*, 2003; Busatto *et al*, 2004) have failed to replicate the multitude of small, clinic-based case-control studies that preceded them (Shenton *et al*, 2001) suggests that the findings of the latter may owe much to the processes involved in selecting cases and controls.

The Strengthening the Reporting of Observational studies in Epidemiology (STROBE) initiative is an attempt to bring about improvements to the methodology and reporting of observational studies, by publishing a checklist with which it is intended all observational research reports will have to comply as a condition of publication (Altman *et al*, 2005). We are optimistic that efforts such as this will improve the standard of reporting and methodology in psychiatric case-control studies in future years.

Although the main aim of our review was to assess potential sources of bias in case-control studies, we noted that many studies had very small sample sizes, with a quarter of all studies having no more than 15 cases. Small sample sizes lead to type 2 error – when a genuine difference between groups is not detected. We also noted that sample sizes varied to a large extent according to the parameter under study. Neuroimaging and 'biological' studies generally had much smaller sample sizes than did genetic and 'pencil and paper' studies. It is difficult to make a general recommendation about the sample size required for the question under study, and variation between methods may be owing to differences in what investigators perceive to be an effect size worth detecting. Differences may also arise because the parameter under study may be measured as a continuous variable (e.g. the volume of a brain structure) or a categorical variable (e.g. the presence of a specific genotype); the use of continuous variables improves power, and therefore smaller sample sizes can be used. However, we also suspect that the expense of performing complex neuroimaging studies or biological assays might mean that these studies are particularly prone to be underpowered.

We were surprised that many studies were individually matched without it being clear that a matched analysis was executed, as this practice results in the needless loss of statistical power (Miettinen, 1970). This and the prevalence of non-equal group sizes in 'matched' studies illustrate some of the many problems with individual matching and explain why this technique has largely been superseded in epidemiology by the use of the more flexible multivariable statistical methods (Prentice, 1976; Rosner & Hennekens, 1978).

This review has several limitations. We undertook to examine studies published only in the highest-impact general psychiatric journals; this was done over a limited period; we only examined one case group and one control group from each study, and the rating scales were simply constructed. We chose the journals with high impact factors to target studies likely to represent accepted practice, where one might expect only examples of good methodology to be accepted, and therefore papers published in less prestigious journals may have even poorer reporting of methodology. The 2-year period we chose was the most recent period for which we had impact factors when the hand-searching was started. We only chose one case group and one control group from each study to simplify our method and analyses. We believe this made little difference to our findings, as most of the studies had only two groups, and in studies with more the methods of selection and reporting of the other groups tended to be similar. Our sampling frame was explicit and representative, including journals from the UK and the USA, and our inclusion and exclusion criteria were predetermined. We feel that the results of this review are likely to represent the standard of global English-language accepted practice of the reporting of psychiatric case-control studies in 2001 and 2002, and we suspect that the standards of reporting of case-control studies are unlikely to have improved markedly since then. The construction of the three rating scales, simply adding the number of questions answered to indicate good practice within the three sections of the questionnaire, was chosen as the most straightforward method of indicating the general quality of the studies. The authors believe that although equating the methodological characteristics of the papers may seem arbitrary, all the items on the questionnaire are important, so none should be deemed less important than any

other. The number of questions in each of the rating scales was small (9, 6 and 2 respectively) which could leave the results vulnerable to floor and ceiling effects, potentially not detecting true associations. Although the numbers are small, on inspection of the data (see Figs 1–3) the authors do not think that large effects are likely to have been undetected.

We have shown that there is a tendency for psychiatric researchers to ignore the potential impact of bias on their results. It is impossible to determine whether the studies we included simply reported their methods inadequately or used inadequate methods. We suggest that researchers have a responsibility to reassure readers that appropriate steps have been taken to eliminate bias, and at present this is not happening.

REFERENCES

- Altman, D. G. (2003)** *Practical Statistics for Medical Research*. Chapman & Hall.
- Altman, D. G., Schulz, K. F., Moher, D., et al (2001)** The revised CONSORT statement for reporting randomized trials: explanation and elaboration. *Annals of Internal Medicine*, **134**, 663–694.
- Altman, D. G., Egger, M., Pocock, S. J., et al (2005)** STROBE Checklist, Version 2. STROBE Initiative. <http://www.pebita.ch/downloadSTROBE/STROBE-Checklist-Version2.pdf>.
- Busatto, G. F., Schaufelberger, M., Perico, C. A. M., et al (2004)** A population-based MRI study of first-episode psychosis in Brazil. *Schizophrenia Research*, **67**, 94.
- Dazzan, P., Morgan, K. D., Suckling, J., et al (2003)** Grey and white matter changes in the A&SOP first-onset psychosis study: a voxel-based analysis of brain structure. *Schizophrenia Research*, **60**, 192.
- Hotopf, M., Lewis, G. & Normand, C. (1997)** Putting trials on trial—the costs and consequences of small trials in depression: a systematic review of methodology. *Journal of Epidemiology and Community Health*, **51**, 354–358.
- Lewis, G. & Pelosi, A. J. (1990)** The case–control study in psychiatry. *British Journal of Psychiatry*, **157**, 197–207.
- Mayes, L. C., Horwitz, R. I. & Feinstein, A. R. (1988)** A collection of 56 topics with contradictory results in case-control research. *International Journal of Epidemiology*, **17**, 680–685.
- Miettinen, O. S. (1970)** Matching and design efficiency in retrospective studies. *American Journal of Epidemiology*, **91**, 111–118.
- Ogundipe, L. O., Boardman, A. P. & Masterson, A. (1999)** Randomisation in clinical trials. *British Journal of Psychiatry*, **175**, 581–584.
- Pocock, S. J., Collier, T. J., Dandreo, K. J., et al (2004)** Issues in the reporting of epidemiological studies: a survey of recent practice. *BMJ*, **329**, 883.
- Prentice, R. (1976)** Use of the logistic model in retrospective studies. *Biometrics*, **32**, 599–606.
- Rosner, B. & Hennekens, C. H. (1978)** Analytic methods in matched pair epidemiological studies. *International Journal of Epidemiology*, **7**, 367–372.
- Sackett, D. L. (1979)** Bias in analytic research. *Journal of Chronic Diseases*, **32**, 51–63.
- Schulz, K. F. (1995a)** *The Methodologic Quality of Randomization as Assessed from Reports of Trials in Specialist and General Medical Journals*. American Association for the Advancement of Science.
- Schulz, K. F. (1995b)** Subverting randomization in controlled trials. *JAMA*, **274**, 1456–1458.
- Schulz, K. F. (1995c)** Unbiased research and the human spirit: the challenges of randomized controlled trials. *Canadian Medical Association Journal*, **153**, 783–786.
- Schulz, K. F. (1996)** Randomised trials, human nature, and reporting guidelines. *Lancet*, **348**, 596–598.
- Shenton, M. E., Dickey, C. C., Frumin, M., et al (2001)** A review of MRI findings in schizophrenia. *Schizophrenia Research*, **49**, 1–52.
- von Elm, E. & Egger, M. (2004)** The scandal of poor epidemiological research. *BMJ*, **329**, 868–869.
- Wacholder, S. (1995)** Design issues in case–control studies. *Statistical Methods in Medical Research*, **4**, 293–309.

Correspondence: Professor Matthew Hotopf, Department of Psychological Medicine, King's College London, Institute of Psychiatry, Weston Education Centre, 10 Cutcombe Rd, London SE5 9RJ, UK. Email: m.hotopf@iop.kcl.ac.uk

(First received 8 June 2006, accepted 1 September 2006)