

DEBATE

What makes a good case–control study?

Design issues for complex traits such as endometriosis

Krina T.Zondervan^{1,3}, Lon R.Cardon¹ and Stephen H.Kennedy²

¹Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford OX3 6BN and ²Nuffield Department of Obstetrics and Gynaecology, University of Oxford, Oxford OX3 9DU, UK

³To whom correspondence should be addressed. E-mail: krina.zondervan@well.ox.ac.uk

The combined investigation of environmental and genetic risk-factors in complex traits will refocus attention on the case–control study. Endometriosis is an example of a complex trait for which most case–control studies have not followed the basic criteria of epidemiological study design. Appropriate control selection has been a particular problem. This article reviews the principles underlying the design of case–control studies, and their application to the study of endometriosis. Only if it is designed well is the case–control study a suitable alternative to the prospective cohort study. Use of newly diagnosed over prevalent cases is preferable, as the latter may alter risk estimates and complicate the interpretation of findings. Controls should be selected from the source population from which cases arose. Potential confounding should be addressed both in studies of environmental and genetic factors. For endometriosis, a possible design would be to: (i) use newly diagnosed cases with ‘endometriotic’ disease; (ii) collect information predating symptom onset; and (iii) use at least one population-based female control group matched on unadjustable confounders and screened for pelvic symptoms. In conclusion, future studies of complex traits such as endometriosis will have to incorporate both environmental and genetic factors. Only adequately designed studies will allow reliable results to be obtained and any true aetiological heterogeneity expected to underlie a complex trait to be detected.

Key words: case–control study/complex trait/endometriosis/risk-factor/study design

Introduction

Investigating the aetiology of complex traits represents a major challenge. The multiple genetic and environmental factors they are caused by are likely to have only modest effect sizes that will vary across populations (Cardon and Bell, 2001). The need to incorporate environmental factors in analyses has refocused attention on traditional epidemiological study designs such as the case–control study (Clayton and McKeigue, 2001). In genetic research settings, concerns for analytical problems such as confounding by population of origin (population stratification) have in the past brought this type of study into disrepute. In epidemiological research, however, similar problems of confounding are regarded as mostly related to poor study design in terms of case and control selection.

Endometriosis is an example of a complex trait with several additional features complicating epidemiological study design, such as the lack of consensus about its precise definition and the need for an invasive procedure to establish the diagnosis. The condition is broadly defined as the presence of endometrial-like tissue outside the uterine cavity, associated with symptoms of dysmenorrhoea, dyspareunia, chronic pelvic pain and sub-fertility. It can only be diagnosed with certainty on histological examination. Disease severity has traditionally been classi-

fied using the revised American Fertility Society (rAFS) system into four stages (minimal to severe) on the basis of observed implant size, presence of cysts and adhesion formation (American Fertility Society, 1985). However, minimal or mild endometriosis is increasingly viewed as part of a normal physiological process, whereas the more severe forms—ovarian cysts and deeply infiltrating lesions—are considered ‘endometriotic disease’ (Koninckx *et al.*, 1999). Thus, endometriosis can be seen as a continuum that is only considered pathological when a certain threshold of severity has been reached. This is similar to many other conditions that are regarded as quantitative traits with a threshold of clinical relevance, such as obesity or various psychological disorders.

Because of the many difficulties inherent in the epidemiological study of endometriosis, Holt and Weiss recently published some excellent recommendations for study design (Holt and Weiss, 2000). They stressed the importance of using a standard definition, and discussed the implications of selecting cases from various source populations. We wish to build on their recommendations by demonstrating how the principles of a well-designed case–control study can be applied in the investigation of both genetic and environmental risk-factors for endometriosis. We note that most of the points raised are

not confined to the study of endometriosis, but are important in case–control studies of any complex trait.

Environmental and genetic epidemiology of endometriosis: research to date

Because of the need for a surgical diagnosis, the prevalence of endometriosis in the general population is unknown. Estimates from asymptomatic fertile subpopulations undergoing tubal ligation have varied greatly, from 0.7 to 43% around a mean of 4% (Eskenazi and Warner, 1997). However, up to 90% of these women were diagnosed with minimal or mild endometriosis.

The main aetiological hypothesis for endometriosis is retrograde menstruation (Sampson, 1927). However, retrograde menstruation has been observed in up to 90% of women (Halme *et al.*, 1984), which implies that other factors must also be involved. Inevitably, the need for a surgical diagnosis has limited studies investigating risk-factors for endometriosis, since they have to be based on selected patient samples.

The evidence that endometriosis is a complex trait is highly suggestive. Reviews of environmental risk-factors, researched independently from genetic factors, have implicated prolonged and heavy menstruation and increased exposure to estrogen (Mangtani and Booth, 1993; Eskenazi and Warner, 1997). Many of these studies failed to take account of basic epidemiological principles in their design. Of 100 studies of environmental risk-factors reviewed by Eskenazi and Warner, only six met the following basic criteria for adequate study design: (i) cohort or case–control design; (ii) surgically confirmed cases; (iii) clearly described criteria for control selection; and (iv) adjustment for confounding factors in the analysis (Eskenazi and Warner, 1997). In a search for studies published since then, we have found only two more studies that conformed to these criteria (Signorello *et al.*, 1997; Pauwels *et al.*, 2001). The total of eight studies, seven of which were of case–control design, varied widely in terms of case definition and control selection (Table I). Apart from generally consistent associations with increasing age and prolonged menstruation, other findings such as for smoking, exercise, body mass index, parity and tampon use were either inconsistent or simply not tested in more than one study (Eskenazi and Warner, 1997). Exposure to 2,3,7,8-tetrachlorodibenzo-p-dioxin has been implicated in primate studies (Rier *et al.*, 1993), but evidence for a role in human endometriosis is limited (Mayani *et al.*, 1997; Pauwels *et al.*, 2001).

Two recent reviews have discussed the evidence for a genetic aetiology of endometriosis (Bischoff and Simpson, 2000; Zondervan *et al.*, 2001a). Genetic factors were implicated by a large twin study, in which 51% of the variance of susceptibility to endometriosis was attributed to genes (Treloar *et al.*, 1999), and by four case–control studies showing that the first-degree relatives of affected women are at 3–9 times increased risk of developing the disease compared with first-degree relatives of controls (Simpson *et al.*, 1980; Lamb *et al.*, 1986; Coxhead and Thomas, 1993; Moen and Magnus, 1993). There have been 11 case–control studies (Table II) that have assessed the influence of specific genetic variants ('functional'

candidate genes), mainly focusing on genes involved in detoxification (*GSTM1, GSTT1, NAT2*), galactose metabolism (*GALT*), differential expression of hormone receptors (*ESR1*) and immunological dysfunction (*IL-1β*). Case–control studies of genetic variants have been substantially smaller than those of environmental factors, and generally lack the power required to detect the moderate effect sizes likely to apply to complex traits such as endometriosis. Most did not comply with the criteria of basic epidemiological study design described above, in particular that of appropriate control selection and adjustment for potential confounders (Zondervan *et al.*, 2001a). This may be due to a misconception that these principles have been developed specifically for studies of environmental factors, which are often difficult to measure, change over time, and the collection of which may be subject to information bias. Nevertheless, as will be discussed in this paper, the choice of case and control selection can also have a profound effect on the results of candidate gene studies.

In the following few paragraphs we briefly discuss the principles of the case–control study and the reasons why appropriate selection of cases, and in particular of controls, is so important.

Principal aims of a case–control study

A case–control study aims to derive a risk estimate for a particular factor of exposure (environmental or genetic) that is as close as possible to the estimate that would have been derived had a prospective cohort study been performed. Cohort studies (in which two or more groups of people free of the disease of interest but different in terms of exposures are followed to investigate who develops the disease) are the 'gold standard' for risk-factor analysis, because they allow the collection of unbiased risk-factor information. When unfeasible, the case–control study (in which exposures are compared between groups of people with and without the disease of interest) can be a good alternative, provided cases and controls are selected appropriately.

In a cohort study, a population consisting of exposed and unexposed individuals is followed for an amount of time, and the incidence rates of disease are compared between the two groups (Greenland and Rothman, 1998).

$$I_{exp} = A_{exp}/T_{exp}$$

$$I_{unexp} = A_{unexp}/T_{unexp}$$

Where I = incidence rate, A = number of affecteds and T = amount of person-time spent in the exposed or unexposed group.

The incidence rate ratio (IR) is then calculated to determine how far from unity (no effect of the risk-factor) this ratio is:

$$IR = I_{exp}/I_{unexp} = A_{exp} \times T_{unexp} / A_{unexp} \times T_{exp}$$

In order for a case–control study to arrive at the same risk estimate as a cohort study, cases should be the same individuals who would have been considered cases in a hypothetical cohort study. Provided this rule is applied, direct estimates for A_{exp} and A_{unexp} can be obtained from a case–control study by

Table I. The eight studies of environmental risk-factors for endometriosis that were considered of adequate epidemiological design. Adapted from Eskenazi and Warner (1997).

Study design	Disease stage ^a	References	Country
Cohort			
17 302 married, Caucasian women, identified from family planning clinics; cases from hospital admission records.	Unknown	Vessey <i>et al.</i> (1993)	UK
Case-control			
286 nulliparous, infertile cases; 3794 controls admitted for delivery (no laparoscopy). Multicentre study.	Unknown	Cramer <i>et al.</i> (1986); Grodstein <i>et al.</i> (1993; 1994)	USA
170 cases; 170 age-matched controls (laparoscopy for other reasons). Record-based study.	Unknown	Makhlouf-Obermeyer <i>et al.</i> (1986)	Lebanon
376 cases; 522 controls admitted for acute conditions (no laparoscopy).	Unknown	Candiani <i>et al.</i> (1991); Parazzini <i>et al.</i> (1994; 1995)	Italy
104 cases; 100 'friend' controls (no laparoscopy)/98 clinic controls (laparoscopy).	I: 34; II: 20; III: 26; IV: 20%	Darrow <i>et al.</i> (1993; 1994); McCann <i>et al.</i> (1993)	USA
126 multiparous cases; 504 controls (both groups laparoscopic sterilization).	I: 91; II: 5; III: 4%	Sangi-Haghpeykar and Poindexter (1995)	USA
50 infertile cases; 89 fertile controls (laparoscopic tubal ligation) and 47 infertile controls (laparoscopy).	Unknown	Signorello <i>et al.</i> (1997)	USA
34 prospectively identified infertile cases; 27 infertile controls (laparoscopy).	I: 50; II: 16; III: 24; IV: 10%	Pauwels <i>et al.</i> (2001)	Belgium

Criteria for adequate epidemiological design were: (i) study design was either cohort or case-control; (ii) cases were surgically confirmed; (iii) criteria for control selection were clearly stated; (iv) adjustment for potential confounders in analyses.

^aAccording to revised AFS classification (1985): I = minimal; II = mild; III = moderate; IV = severe.

measuring exposure status in the cases. The controls must be selected in such a way that allows estimation of T_{unexp}/T_{exp} . As long as controls are sampled from the same population from which the cases arose, and the sampling is independent of exposure status, the ratio of unexposed versus exposed controls (U_{unexp}/U_{exp}) is the same as the ratio of unexposed versus exposed person-time (T_{unexp}/T_{exp}) in that population. In other words, controls should reflect the exposure distribution of the population that cases were sampled from. Substituting T_{unexp}/T_{exp} with U_{unexp}/U_{exp} gives the familiar equation for the odds ratio (OR), the effect measure derived from a case-control study:

$$OR = A_{exp} \times U_{unexp} / A_{unexp} \times U_{exp}$$

Where A = number of affecteds (cases) and U = number of unaffecteds (controls).

The above translates into the following general principles for designing a case-control study (Rothman and Greenland, 1998a). Firstly, cases should be incident (newly arising) cases that are recruited prospectively from a certain population during the time period of the study, and for whom risk-factor information is collected retrospectively. Secondly, as each new case arises, one or more controls are sampled from the same population and their risk-factor information is collected. Every time a control is selected, he/she is not removed from the sampling population but remains eligible for future sampling either as a control or case. This means that theoretically, an unaffected individual could be selected more than once as a control, and could subsequently be selected as a case if he/she develops the disease later in the study period (although in practice the likelihood of this scenario is small).

Practical considerations when investigating endometriosis

Incident versus prevalent cases

The main problem in applying the above guidelines often lies with the identification of incident cases. Endometriosis is no exception. Not only is there a substantial delay between symptom onset and diagnosis (Hadfield *et al.*, 1996), we do not know what pathological changes qualify as 'onset' of disease, nor do we have the means to measure these changes as they occur. For most non-infectious conditions, the term 'incident' is rather an artificial concept. Disease onset is more appropriately viewed as a continuum of biological changes which, once a certain threshold has been reached, is considered clinically relevant and termed 'disease'.

In practice, newly diagnosed cases are usually taken to be 'incident'. Their use is highly preferable over prevalent cases (i.e. all those in a population having the disease at a specific point in time irrespective of time since diagnosis) because they minimize the chance that an observed effect of an environmental risk-factor is the result of diagnosis. One can easily imagine this for behavioural risk-factors: a person may be more likely to stop smoking or change exercise frequency because of a certain diagnosis. Equally, a person may change habits because of the onset of symptoms, a time-point which could pre-date diagnosis considerably. In data collection every effort must therefore be taken to determine environmental exposures prior to the onset of disease symptoms.

Even the use of newly diagnosed cases with risk-factor information pre-dating symptom onset cannot circumvent the problem that certain exposures may have changed as a result of subclinical disease. For example, heavy menstrual bleeding that pre-dated symptom onset in endometriosis may imply a

Table II. The 11 case-control studies of genetic risk-factors for endometriosis

Study design (polymorphism)	Candidate gene	Disease stage ^a	References	Country
42 cases; 67 male + female population controls from north-west and 157 from central-eastern region.	<i>GSTM1</i> (null allele)	Unknown %	Baranov <i>et al.</i> (1996)	Russia
33 infertile cases; 111 pre-menopausal, female, population controls.	<i>GALT</i> (N314D)	I: 45%; II: 21%; III: 21%; IV: 12%	Cramer <i>et al.</i> (1996)	USA
65 cases; 72 female controls admitted for voluntary abortion.	<i>GSTM1</i> (null allele); <i>GSTT1</i> (null allele); <i>NAT2</i> (*5, *6, *7 alleles)	I/II: 55 III/IV: 45%	Baranova <i>et al.</i> (1997; 1999)	France
78 newly diagnosed cases; 248 healthy female volunteers from same region.	<i>GALT</i> (N314D; Q188R)	*Ovarian endometriosis*	Morland <i>et al.</i> (1998)	UK
57 familial (<i>FH+</i>) and 91 sporadic (<i>FH-</i>) cases; 95 male population controls;	<i>GALT</i> (N314D); <i>GSTM1</i> (null allele);	<i>FH+</i> III/IV: 100%; <i>FH-</i> I/II: 88%; III/IV: 12%	Hadfield <i>et al.</i> (1999; 2001)	UK
53 pre-menopausal women aged 40-50 years with normal pelvis at hysterectomy.	<i>GSTT1</i> (null allele); <i>CYP11A1</i> (MspI)			
57 cases aged 24-38; 57 female controls from same region, matched on age and social status ^b .	<i>ESR1</i> (Pvu II); [TA]n	I: 49%; II: 35%; III: 12; IV: 4%	Georgiou <i>et al.</i> (1999)	Greece
109 cases; 179 female population controls.	<i>ESR1</i> (Pvu II)	I/II: 22; III/IV: 78%	Kitawaki <i>et al.</i> (2001)	Japan
84 newly diagnosed cases; 219 healthy female volunteers from same region.	<i>GSTM1</i> (null allele)	*Ovarian endometriosis*	Baxter <i>et al.</i> (2001)	UK
54 familial (<i>FH+</i>); 95 male population controls; 24 pre-menopausal women aged 40-50 years with normal pelvis at hysterectomy.	<i>NAT2</i> (*5, *6, *7 alleles)	III/IV: 100%	Nakago <i>et al.</i> (2001)	UK
85 unrelated cases selected from 205 with a family history nation-wide; 213 female + male volunteers.	<i>GALT</i> (N314D)	I/II: 6%; III/IV: 94%	Stefansson <i>et al.</i> (2001)	Iceland
102 pre-menopausal surgically diagnosed cases; 119 controls 'without endometriosis'.	<i>P21</i> : cdn 31 (Ser/Arg); <i>IL-1β</i> : 511 prom. (C-T); <i>IL-1β</i> : exon 5 (E1/E2); <i>IL-1Ra</i> (I/II/III/IV)	Unknown	Hsieh <i>et al.</i> (2001b; 2001a)	Taiwan

^aAccording to revised AFS classification (1985); I = minimal; II = moderate; III = mild; IV = severe.

^bThe single study that adjusted for any potential confounding factors (through matching).

role in its causation, but could also be a result of early physiological changes associated with the condition. Such possible explanations should be borne in mind when interpreting the study results.

When prevalent rather than incident cases are used, an additional problem is the validity of the resulting effect size, the OR. In a case-control design which uses prevalent cases, it is impossible to sample controls as cases arise. Instead, they are sampled from the people that are unaffected in the population, after affecteds have been excluded. The OR is then approximated by:

$$OR = \frac{A_{exp} \times (N_{unexp} - A_{unexp})}{A_{unexp} \times (N_{exp} - A_{exp})}$$

(Rothman and Greenland, 1998a)

Where A = number of affecteds and N = number of individuals in the population from which affecteds are sampled.

In this situation, the OR only provides a reasonable approximation of the relative risk that would have been obtained in a cohort study if the disease is relatively rare (general rule of thumb: prevalence A/N <20%). For diseases more common than this, an increase in risk associated with an exposure will produce inflated ORs (Kirkwood, 1988).

The main advantage of studying candidate genes rather than environmental factors is that the exposure of interest usually remains constant (except for situations in which differential expression at different times in life occurs, or somatic mutations influence outcome). Genetic factors cannot be a result of symptom onset or diagnosis. However, many of the concerns raised also apply to studies of genetic risk-factors. Every effort must be taken for controls to be sampled from the source population from which cases arose. Choosing to use incident or prevalent cases also has certain consequences. Sampling of prevalent cases will provide a mixture that is skewed towards individuals who have had the condition longer (Freeman and Hutchison, 1980). Different genetic factors may be found in studies using prevalent cases than in those using incident cases: the former could be more important for disease maintenance, whereas the latter could relate more to the onset of disease. Prior hypotheses for the disease model could help in the choice of incident versus prevalent cases, whereas using duration of the disease as a co-variate in the analysis may also be of benefit.

Case definition in endometriosis research

As shown in Tables I and II, many different definitions of endometriosis have been used, most of which appear to be based on opportunistic groups of prevalent cases that were seen in the study clinic at the time. Holt and Weiss rightly commented that for study results to become comparable, a standard definition has to be used (Holt and Weiss, 2000). The rAFS classification system (American Fertility Society, 1985) is not particularly suited to this purpose, as it was principally designed to categorize women according to the probability of conceiving and does not correlate well with pelvic pain symptoms (Porpora *et al.*, 1999). Because of the high frequency with which minimal/mild endometriosis is found in asymptomatic women, and current theories of these disease stages representing a normal physiological process, it appears logical to limit case definition to more severe stages. Holt and

Weiss proposed a standard definition of definite and possible 'endometriotic disease' based on a combination of surgical observations and pelvic symptoms (Table III). In addition, a recent study showed that ovarian endometriosis, often present in endometriotic disease, can be accurately diagnosed through high-resolution ultrasound (Eskenazi *et al.*, 2001).

Control selection

The choice of a control group is entirely determined by the definition and selection of the case group. More specifically, the source population from which controls are sampled should be that from which cases are also sampled. This is true for studies of environmental as well as genetic factors. An appropriate choice of controls will allow the allele frequencies of cases to be compared with those of their source population, and thus minimize the chance of population stratification (finding spurious associations).

A case-control study can be restricted to any (sub)type of case that may be of interest, as long as controls are selected appropriately for these case groups. Of course, the more restricted a definition of a case in terms of subtype or setting, the more difficult it becomes to identify the population from which such cases arose. It is important that controls should have had the same opportunity to develop the disease of interest, and—had they done so—they would have had the same opportunity as cases to have been included in the study. By definition, this means they should be women, since men cannot express the phenotype. Some studies of candidate genes have used males as controls, given that their allele frequencies were known to represent those in the source population (Hadfield *et al.*, 1999; Nakago *et al.*, 2001; Stefansson *et al.*, 2001). However, using male rather than female controls for whom disease status is unknown has no added benefit: had they been women, they may have been a case. Moreover, the use of male controls may become a potential source of bias when environmental exposures are also included in the study, as many exposure profiles are likely to differ from those among women, thus producing spurious associations.

The choice of control groups in endometriosis research has focused on the concern that controls should be free of disease. Because of the requirement for a surgical diagnosis, these control groups have included fertile women undergoing laparoscopic sterilization and women who underwent laparoscopy for infertility unrelated to endometriosis. It is unlikely that these groups were representative of the source population from which cases were derived.

Rather than selecting controls who underwent laparoscopy, it would be advantageous to find a control group that is more population-based. The main concern against this option would be that such a control group could contain a substantial number of undiagnosed cases, thereby diluting the risk factor effects. This concern is likely to be unjustified. In a community survey, Zondervan *et al.* found a prevalence of 24.0% for chronic pelvic pain in the UK (Zondervan *et al.*, 2001b). A total of 28.3% reported having had chronic pelvic pain, infertility (defined as inability to conceive for at least 12 months) or both (unpublished results). In their review, Eskenazi and Warner noted that in studies of women who underwent laparoscopy for

Table III. Standard definition proposed by Holt and Weiss (2000) of endometriotic disease

Disease status	Signs and symptoms
Definite disease	One or more of the following: Ovarian endometriomas of any size Pelvic endometriotic implants of any size >5 mm deep Pelvic endometriotic implants of any size with adhesions not attributable to another cause
Possible disease	Pelvic endometriotic implants not meeting the above criteria, plus one or more of the following: Infertility ^a Moderate to severe dysmenorrhoea ^b Moderate to severe dyspareunia ^c Moderate to severe pelvic pain ^d

^a12 months or more of unprotected sexual intercourse without conception, not attributable to another cause.
^bModerate dysmenorrhoea = occasional loss of work from dysmenorrhoea; severe dysmenorrhoea = incapacitation resulting in the necessity to stay in bed for ≥1 days.
^cModerate dyspareunia = intercourse painful to the point of interruption; severe dyspareunia = avoidance of intercourse because of pelvic pain.
^dModerate pelvic pain = noticeable pelvic discomfort for most of the menstrual cycle; severe pelvic pain = pain persisting during the cycle or requiring strong analgesics.

pelvic pain or infertility, the prevalence of endometriosis was 20% on average, and that around a third of these had moderate/severe disease (Eskenazi and Warner, 1997). From these findings, it appears that the community prevalence of more severe stages of endometriosis is probably <2%. Community-based control groups are therefore unlikely to contain many undiagnosed cases, especially if they are screened for moderate to severe pelvic symptoms. Therefore, the inclusion of undiagnosed cases will dilute the observed effects of risk-factors only to a marginal extent. Instead, it appears much more beneficial to focus research efforts on defining a control population that is representative of the source population of cases, than to be overly concerned about obtaining a completely disease-free control group.

Community-based controls can be a random sample of a particular population, or more selected groups such as neighbourhood and friend controls. Neighbourhood controls are subjects who are sampled from the same neighbourhood as cases arose from (Rothman and Greenland, 1998a). However, if any of the exposures of interest are related to the living environment of cases, using neighbourhood controls will prevent this factor from being identified as a risk-factor (over-matching). Using friend controls could cause similar problems, as friends may be more similar to cases in certain behavioural exposures. In addition, friend controls are identified by cases themselves and are therefore less likely to be chosen independent of their exposure status (thus potentially causing bias).

An alternative strategy is to use hospital-based controls diagnosed with unrelated conditions. The main hypothesis behind this approach is that such controls are better matched to cases on various potential confounding factors that are impossible to measure, such as referral or health care seeking pattern and socio-economic status. A potential problem is that they are not selected at random from the source population of cases and may thus be unrepresentative of the exposure distribution in that population. Using several different diagnostic groups can dilute the biasing effects of including a specific diagnostic group that is unrepresentative of the

source population. Suggestions for appropriate groups very much depend on the health care system of the country under study. Essentially, controls should have had the same opportunity and inclination to attend for their respective diagnoses as the case group did for theirs. For endometriosis, an example could be women attending for treatment of chronic, non-cancerous conditions such as asthma.

Lastly, as there is almost never one ideal control group, an obvious solution is to use multiple groups. However, this makes the study more expensive and the interpretation of the analyses more complicated. If control groups differ in their exposure patterns, it is difficult to find out which one most represents the true exposure distribution of the source population (Rothman and Greenland, 1998a).

Matching

All control groups that are not randomly selected from a population (neighbourhood, friend and hospital-based controls) essentially represent forms of matching. Matching refers to the selection of controls that are as similar as possible to cases with respect to the distribution of one or more potential confounding factors that are difficult to measure. Cases can be matched to controls on an individual basis or in strata of exposure values (frequency matching). Matching is not without disadvantages. Matching factors can no longer be investigated in the analysis, as they need to be used as stratification variables. Furthermore, there are situations in which over-matching can occur (Rothman and Greenland, 1998b). Matching on a variable that is associated only with exposure, not with disease, reduces statistical efficiency: the investigator has to stratify on the matching variable, while in an unmatched design adjustment for the variable would have been unnecessary. Matching on variables that are affected by exposure or disease (such as symptoms) can cause bias and thus affect the validity of the results. In a study of endometriosis, for example, one would never match on level of pelvic pain or parity, as this would make controls more similar to cases for various

potential risk-factors that affect both endometriosis and pain or infertility.

There are various situations in which individual matching could be appropriate in the study of endometriosis. As endometriosis is an age-related condition (Eskenazi and Warner, 1997) and age is also related to many exposures, it is an important confounder. It can be controlled for in the analysis, but it may be statistically more efficient to match on age. If there is concern for bias due to differential access to health care, various forms of matching could be used. For example, in countries where patients have to be referred from a primary care doctor to a gynaecological clinic, controls could be matched to cases on the particular primary care doctor that referred the case. In multicentre studies, one would also match on centre.

A special form of matching, developed in genetic research, is the use of family-based controls such as siblings or cousins. The main reason for its development was to avoid confounding in comparing allele frequencies between cases and controls by population origin (population stratification), which was thought to be a major concern when using population-based controls. Although family-based controls are appropriate when studying the effect of genes only, they become a problem when wishing to incorporate environmental factors in the analysis. The main disadvantage is over-matching on environmental factors. For example, siblings are more likely to share environmental exposures because they have been brought up in the same surroundings. This means that the number of discordant case-control units is smaller and therefore many more are needed for the study to have sufficient power. The power is further decreased by the fact that a stratified analysis has to be performed to allow for the matching on family. Lastly, cases may or may not have (a different number of) eligible relatives. This creates the problem of who to select, and how to adjust the analysis to incorporate within-family dependency of the measurements (Weinberg and Umbach, 2000). The latter problem is an important limitation of the design for endometriosis research. As endometriosis is heritable and related to infertility, cases may—on average—have fewer eligible blood relatives than controls without the condition.

Other study designs have been developed to avoid population stratification in genetic studies, such as the case-only design and various case-parent designs (haplotype relative risk, transmission disequilibrium test, pseudo-sib) that use non-transmitted alleles from parents in various ways to create 'controls'. The relative power of these various types of association studies as opposed to linkage studies has been well described within the field of statistical genetics (Risch and Merikangas, 1996; Risch and Teng, 1998; Teng and Risch, 1998). Weinberg and Umbach have also provided a detailed comparison between such family-based methods of association and population-based case-control designs, concluding that most family-based methods were unsuitable for the estimation of the main effects of genes and exposures as well as their interaction (Weinberg and Umbach, 2000). A method worth mentioning that tries to address the potential problem of population stratification in standard case-control studies, is that of 'genomic control'. This tests the presence of population

stratification in a standard case-control study by comparing allelic frequencies of randomly selected anonymous genes between cases and controls (Bacanu *et al.*, 2000; Pritchard *et al.*, 2000). If these frequencies differ systematically then population stratification is likely and should be adjusted for by using stratified analyses. This approach could be suitable for the study of endometriosis. However, although the power of the method may already be reasonable when using 20 markers, its application may at present be limited because of genotyping cost in the large case-control studies needed to study complex traits.

The study of gene-environment interaction

Recent debate has somewhat questioned the scientific merit of studying statistical interactions between genes and environment in complex diseases (Clayton and McKeigue, 2001). Nevertheless, it also suggested that the population-based case-control study is the most suited to such investigations. However, an important limiting factor in studying interaction is sufficient power. Smith and Day calculated that to detect an interaction between two dichotomous variables with main effects of similar magnitude in an unmatched case-control study, study size would have to be increased at least 4-fold (Smith and Day, 1984).

Matching and counter-matching can provide ways to improve the power for studying gene-environment interaction, but only in situations where the exposures are rare (Smith and Day, 1984; Andrieu *et al.*, 2001). These methods also assume that the gene-environment interaction of interest is known prior to the design of the study, whereas in practice, investigators tend to look only for interaction between factors identified from the study with main effects that are large enough to be of interest. This approach does not allow for factors that have small main effects, but for which the interaction produces higher risk estimates. This is more likely to be the scenario in the investigation of complex traits including endometriosis.

Conclusions

Appropriate case definition and control selection is vital in determining the validity and reproducibility of case-control studies of complex traits. With respect to endometriosis, insufficient attention has been paid to this topic, especially in studies investigating the effects of candidate genes.

Future studies of complex traits will increasingly have to incorporate both environmental and genetic factors. Since individual effect sizes for risk-factors underlying complex traits are unlikely to be large, the collection of accurate, unbiased and comparable data from sufficiently large samples will be of the utmost importance. It is only if designs of studies in different populations are valid and consistent that we will be able to compare their results and differentiate between true aetiologic heterogeneity expected to underlie a complex trait and effects due to design differences and inadequacies. In view of the generally poor study designs to date, this appears of particular relevance to endometriosis.

Acknowledgement

K.T.Z. is supported by an MRC Special Training Fellowship in Bioinformatics.

References

- American Fertility Society (1985) Revised American Fertility Society classification of endometriosis. *Fertil. Steril.*, **43**, 351–352.
- Andrieu, N., Goldstein, A.M., Thomas, D.C. and Langholz, B. (2001) Counter-matching in studies of gene–environment interaction: efficiency and feasibility. *Am. J. Epidemiol.*, **153**, 265–274.
- Bacanu, S., Devlin, B. and Roeder, K. (2000) The power of genomic control. *Am. J. Hum. Genet.*, **66**, 1933–1944.
- Baranov, V.S., Ivaschenko, T., Bakay, B., Aseev, M., Belotserkovskaya, R., Baranova, H., Malet, P., Perriot, J., Mouraire, P., Baskakov, V.N. *et al.* (1996) Proportion of the GSTM1 0/0 genotype in some Slavic populations and its correlation with cystic fibrosis and some multifactorial diseases. *Hum. Genet.*, **97**, 516–520.
- Baranova, H., Bothorishvili, R., Canis, M., Albuissou, E., Perriot, S., Glowaczower, E., Bruhat, M.A., Baranov, V. and Malet, P. (1997) Glutathione S-transferase M1 gene polymorphism and susceptibility to endometriosis in a French population. *Mol. Hum. Reprod.*, **3**, 775–780.
- Baranova, H., Canis, M., Ivaschenko, T., Albuissou, E., Bothorishvili, R., Baranov, V., Malet, P. and Bruhat, M.A. (1999) Possible involvement of arylamine N-acetyltransferase 2, glutathione S-transferases M1 and T1 genes in the development of endometriosis. *Mol. Hum. Reprod.*, **5**, 636–641.
- Baxter, S.W., Thomas, E.J. and Campbell, I.G. (2001) GSTM1 null polymorphism and susceptibility to endometriosis and ovarian cancer. *Carcinogenesis*, **22**, 63–65.
- Bischoff, F.Z. and Simpson, J.L. (2000) Heritability and molecular genetic studies of endometriosis. *Hum. Reprod. Update*, **6**, 37–44.
- Candiani, G.B., Danesino, V., Gastaldi, A., Parazzini, F. and Ferraroni, M. (1991) Reproductive and menstrual factors and risk of peritoneal and ovarian endometriosis. *Fertil. Steril.*, **56**, 230–234.
- Cardon, L.R. and Bell, J.I. (2001) Association study designs for complex diseases. *Nat. Rev. Genet.*, **2**, 91–99.
- Clayton, D. and McKeigue, P.M. (2001) Epidemiological methods for studying genes and environmental factors in complex disease. *Lancet*, **358**, 1356–1360.
- Coxhead, D. and Thomas, E.J. (1993) Familial inheritance of endometriosis in a British population. A case control study. *J. Obstet. Gynaecol.*, **13**, 42–44.
- Cramer, D.W., Wilson, E., Stillman, R.J., Berger, M.J., Belisle, S., Schiff, I., Albrecht, B., Gibson, M., Stadel, B.V. and Schoenbaum, S.C. (1986) The relation of endometriosis to menstrual characteristics, smoking and exercise. *JAMA*, **255**, 1904–1908.
- Cramer, D.W., Hornstein, M.D. and Barbieri, R.L. (1996) Endometriosis associated with the N314D mutation of galactose-1-phosphate uridylyl transferase (GALT). *Mol. Hum. Reprod.*, **2**, 149–152.
- Darrow, S.L., Vena, J.E., Batt, R.E., Zielezny, M.A., Michalek, A.M. and Selman, S. (1993) Menstrual cycle characteristics and the risk of endometriosis. *Epidemiology*, **4**, 135–142.
- Darrow, S.L., Selman, S., Batt, R.E., Zielezny, M.A. and Vena, J.E. (1994) Sexual activity, contraception and reproductive factors in predicting endometriosis. *Am. J. Epidemiol.*, **140**, 500–509.
- Eskenazi, B. and Warner, M.L. (1997) Epidemiology of endometriosis. *Obstet. Gynecol. Clin. North Am.*, **24**, 235–258.
- Eskenazi, B., Warner, M., Bonsignore, L., Olive, D., Samuels, S. and Vercellini, P. (2001) Validation study of nonsurgical diagnosis of endometriosis. *Fertil. Steril.*, **76**, 929–935.
- Freeman, J. and Hutchison, G.B. (1980) Prevalence, incidence and duration. *Am. J. Epidemiol.*, **112**, 707–723.
- Georgiou, I., Syrrou, M., Bouba, I., Dalkalitsis, N., Paschopoulos, M., Navrozoglou, I. and Lolis, D. (1999) Association of estrogen receptor gene polymorphisms with endometriosis. *Fertil. Steril.*, **72**, 164–166.
- Greenland, S. and Rothman, K.J. (1998) Measures of effect and measures of association. In *Modern Epidemiology*, 2nd edn. Lippincott-Raven, Philadelphia, USA, pp. 47–64.
- Grodstein, F., Goldman, M.B., Ryan, L. and Cramer, D.W. (1993) Relation of female infertility to consumption of caffeinated beverages. *Am. J. Epidemiol.*, **137**, 1353–1360.
- Grodstein, F., Goldman, M.B. and Cramer, D.W. (1994) Infertility in women and moderate alcohol use [see comments]. *Am. J. Public Health*, **84**, 1429–1432.
- Hadfield, R., Mardon, H., Barlow, D.H. and Kennedy, S.H. (1996) Delay in diagnosis of endometriosis: a survey of women from the USA and the UK. *Hum. Reprod.*, **11**, 878–880.
- Hadfield, R.M., Manek, S., Nakago, S., Mukherjee, S., Weeks, D.E., Mardon, H.J., Barlow, D.H. and Kennedy, S.H. (1999) Absence of a relationship between endometriosis and the N314D polymorphism of galactose-1-phosphate uridylyl transferase in a UK population. *Mol. Hum. Reprod.*, **5**, 990–993.
- Hadfield, R.M., Manek, S., Weeks, D.E., Mardon, H.J., Barlow, D.H., Kennedy, S.H. and OXEGENE Collaborative Group (2001) Linkage and association studies of the relationship between endometriosis and genes encoding the detoxification enzymes GSTM1, GSTT1 and CYP1A1. *Mol. Hum. Reprod.*, **7**, 1073–1078.
- Halme, J., Hammond, M.G., Hulka, J.F., Raj, S.G. and Talbert, L.M. (1984) Retrograde menstruation in healthy women and in patients with endometriosis. *Obstet. Gynecol.*, **64**, 151–154.
- Holt, V.L. and Weiss, N.S. (2000) Recommendations for the design of epidemiologic studies of endometriosis. *Epidemiology*, **11**, 654–659.
- Hsieh, Y.Y., Chang, C.C., Tsai, F.J., Wu, J.Y., Shi, Y.R., Tsai, H.D. and Tsai, C.H. (2001a) Polymorphisms for interleukin-1 beta (IL-1 beta)-511 promoter, IL-1 beta exon 5 and IL-1 receptor antagonist: nonassociation with endometriosis. *J. Assist. Reprod. Genet.*, **18**, 506–511.
- Hsieh, Y.Y., Tsai, F.J., Chang, C.C., Chen, W.C., Tsai, C.H., Tsai, H.D. and Lin, C.C. (2001b) p21 gene codon 31 arginine/serine polymorphism: non-association with endometriosis. *J. Clin. Lab. Anal.*, **15**, 184–187.
- Kirkwood, B.R. (1988) Cohort and case-control studies. In *Essentials of Medical Statistics*. Blackwell Scientific Publications, Oxford, UK, pp. 173–183.
- Kitawaki, J., Obayashi, H., Ishihara, H., Koshiba, H., Kusuki, I., Kado, N., Tsukamoto, K., Hasegawa, G., Nakamura, N. and Honjo, H. (2001) Oestrogen receptor-alpha gene polymorphism is associated with endometriosis, adenomyosis and leiomyomata. *Hum. Reprod.*, **16**, 51–55.
- Koninckx, P.R., Barlow, D. and Kennedy, S. (1999) Implantation versus infiltration: the Sampson versus the endometriotic disease theory. *Gynecol. Obstet. Invest.*, **47** (Suppl. 1), 3–10.
- Lamb, K., Hoffmann, R.G. and Nichols, T.R. (1986) Family trait analysis: a case-control study of 43 women with endometriosis and their best friends. *Am. J. Obstet. Gynecol.*, **154**, 601.
- Makhlouf-Obermeyer, C., Armenian, H.K. and Azoury, R. (1986) Endometriosis in Lebanon. A case-control study. *Am. J. Epidemiol.*, **124**, 762–767.
- Mangtani, P. and Booth, M. (1993) Epidemiology of endometriosis. *J. Epidemiol. Commun. Health*, **47**, 84–88.
- Mayani, A., Barel, S., Soback, S. and Almagor, M. (1997) Dioxin concentrations in women with endometriosis. *Hum. Reprod.*, **12**, 373–375.
- McCann, S.E., Freudenheim, J.L., Darrow, S.L., Batt, R.E. and Zielezny, M.A. (1993) Endometriosis and body fat distribution. *Obstet. Gynecol.*, **82**, 545–549.
- Moen, M.H. and Magnus, P. (1993) The familial risk of endometriosis. *Acta Obstet. Gynecol. Scand.*, **72**, 560–564.
- Morland, S.J., Jiang, X., Hitchcock, A., Thomas, E.J. and Campbell, I.G. (1998) Mutation of galactose-1-phosphate uridylyl transferase and its association with ovarian cancer and endometriosis. *Int. J. Cancer*, **77**, 825–827.
- Nakago, S., Hadfield, R.M., Zondervan, K.T., Mardon, H., Manek, S., Weeks, D.E., Barlow, D.H. and Kennedy, S.H. (2001) Association between endometriosis and N-acetyl Transferase 2 polymorphisms in the UK population. *Mol. Hum. Reprod.*, **7**, 1079–1083.
- Parazzini, F., Ferraroni, M., Bocciolone, L., Tozzi, L., Rubessa, S. and La-Vecchia, C. (1994) Contraceptive methods and risk of pelvic endometriosis. *Contraception*, **49**, 47–55.
- Parazzini, F., Ferraroni, M., Fedele, L., Bocciolone, L., Rubessa, S. and Riccardi, A. (1995) Pelvic endometriosis: reproductive and menstrual risk factors at different stages in Lombardy, northern Italy. *J. Epidemiol. Commun. Health*, **49**, 61–64.
- Pauwels, A., Schepens, P.J., D'Hooghe, T., Delbeke, L., Dhont, M., Brouwer, A. and Weyler, J. (2001) The risk of endometriosis and exposure to dioxins and polychlorinated biphenyls: a case-control study of infertile women. *Hum. Reprod.*, **16**, 2050–2055.
- Porpora, M.G., Koninckx, P.R., Piazzze, J., Natili, M., Colagrande, S. and Cosmi, E.V. (1999) Correlation between endometriosis and pelvic pain. *J. Am. Assoc. Gynecol. Laparosc.*, **6**, 429–434.
- Pritchard, J.K., Stephens, M., Rosenberg, N.A. and Donnelly, P. (2000) Association mapping in structured populations. *Am. J. Hum. Genet.*, **67**, 170–181.

- Rier, S.E., Martin, D.C., Bowman, R.E., Dmowski, W.P. and Becker, J.L. (1993) Endometriosis in rhesus monkeys (*Macaca Mulatta*) following chronic exposure to 2, 3, 7, 8-tetrachlorodibenzo-p-dioxin. *Fund. Appl. Toxicol.*, **21**, 433–441.
- Risch, N. and Merikangas, K. (1996) The future of genetic studies of complex human diseases. *Science*, **273**, 1516–1517.
- Risch, N. and Teng, J. (1998) The relative power of family-based and case-control designs for linkage disequilibrium studies of complex human diseases. I. DNA pooling. *Genome Res.*, **8**, 1273–1288.
- Rothman, K.J. and Greenland, S. (1998a) Case-control studies. In *Modern Epidemiology*, 2nd edn. Lippincott-Raven, Philadelphia, USA, pp. 93–114.
- Rothman, K.J. and Greenland, S. (1998b) Matching. In *Modern Epidemiology*, 2nd edn. Lippincott-Raven, Philadelphia, USA, pp. 147–162.
- Sampson, J.A. (1927) Peritoneal endometriosis due to the menstrual dissemination of endometrial tissue into the peritoneal cavity. *Am. J. Obstet. Gynecol.*, **14**, 469.
- Sangi-Haghpeykar, H. and Poindexter, A.N. (1995) Epidemiology of endometriosis among parous women. *Obstet. Gynecol.*, **85**, 983–992.
- Signorello, L.B., Harlow, B.L., Cramer, D.W., Spiegelman, D. and Hill, J.A. (1997) Epidemiologic determinants of endometriosis: a hospital-based case-control study. *Am. Epidemiol.*, **7**, 267–274.
- Simpson, J.L., Elias, S., Malinak, L.R. and Buttram, V.C.J. (1980) Heritable aspects of endometriosis. I. Genetic studies. *Am. J. Obstet. Gynecol.*, **137**, 327–331.
- Smith, P.G. and Day, N.E. (1984) The design of case-control studies: the influence of confounding and interaction effects. *Int. J. Epidemiol.*, **13**, 356–365.
- Stefansson, H., Einarsdottir, A., Geirsson, R.T., Jonsdottir, K., Sverrisdottir, G., Gudnadottir, V.G., Gunnarsdottir, S., Manolescu, A., Gulcher, J. and Stefansson, K. (2001) Endometriosis is not associated with or linked to the GALT gene. *Fertil. Steril.*, **76**, 1019–1022.
- Teng, J. and Risch, N. (1998) The relative power of family-based and case-control designs for linkage disequilibrium studies of complex human diseases. II. Individual genotyping. *Genome Res.*, **9**, 241.
- Treloar, S.A., O'Connor, D.T., O'Connor, V.M. and Martin, N.G. (1999) Genetic influences of endometriosis in an Australian twin sample. *Fertil. Steril.*, **71**, 701–710.
- Vessey, M.P., Villard-Mackintosh, L. and Painter, R. (1993) Epidemiology of endometriosis in women attending family planning clinics. *Br. Med. J.*, **306**, 182–184.
- Weinberg, C.R. and Umbach, D.M. (2000) Choosing a retrospective design to assess joint genetic and environmental contributions to risk. *Am. J. Epidemiol.*, **152**, 197–203.
- Zondervan, K.T., Cardon, L.R. and Kennedy, S.H. (2001a) The genetic basis of endometriosis. *Curr. Opin. Obstet. Gynecol.*, **13**, 309–314.
- Zondervan, K.T., Yudkin, P.L., Vessey, M.P., Jenkinson, C.P., Dawes, M.G., Barlow, D.H. and Kennedy, S.H. (2001b) The community prevalence of chronic pelvic pain in women and associated illness behaviour. *Br. J. Gen. Pract.*, **51**, 541–547.