

As mentioned in the previous workshop,¹ with a comparison or referent group, a case series can be utilised for studies looking into risk factors or aetiology.

Case-referent studies

A comparison of the past exposure experiences between a case series or group with the disease of interest and a referent group without the index disease is called a case-referent study. This term is preferred over the more commonly used 'case-control study' to avoid confusion with the control (no intervention) group in a randomised controlled trial, as the 'control group' in a 'case-control study' actually refers to a referent group without the index disease, and there is nothing the investigator(s) can 'control' in this type of observational study apart from the selection. A well-conducted case-referent study can provide very useful evidence to support or refute a cause-outcome hypothesis. However, selection bias, information bias, and confounding need to be carefully considered and managed.

Selection bias

As with a case series, the case group with consecutive new cases seen in a hospital or clinic setting that fulfill predetermined inclusion and exclusion criteria should be representative. However, if the case group is recruited from all cases attending a certain clinic or hospital during a certain period of time, it may include follow-up or existing (prevalent) cases in addition to new-onset (incident) cases which can result in bias. Cases that are more aggressive or have a progressive course could have died early and so be less likely to be seen at follow-up. In general, prevalent cases have a better prognosis and examining factors associated with them could result in a mixture of risk factors (leading to occurrence of the disease) and prognostic factors (leading to a better prognosis among the cases), which makes interpretation very difficult. By and large, cases are quite willing to participate and self-selection bias from non-response is usually not a major problem.

A case-referent study also involves a referent group for comparison, for which subject selection is not necessarily straightforward. There are two common approaches—using hospital patients without the index disease (hospital referents) or 'healthy' subjects from the community where the cases come from (community referents). In the later situation, if the cases used are also a representative

sample of all cases in a population, the study can be called a population-based case-referent study. Selecting representative referent subjects from the community/population is usually quite straightforward and depends on standard probability sampling methods, provided that a direct or indirect list of all persons is available (eg telephone directory, household list, list of registered persons in the community). Unfortunately, the response rate is usually disappointing (can be <50%) in many free societies and possible self-selection bias can be a serious concern. Hospital referents usually provide a pretty high response rate (approximately 90% or above), but there are major problems associated with the selection of an appropriate referent group among patients without the index disease. Depending on the risk factor(s) under study, using hospital referents may over- or under-estimate the effect(s). For example, tobacco smoking increases the risk of a range of different diseases, so hospital referents (especially those with cardiopulmonary diseases) are more likely to be smokers than the general population, and examining the effect of smoking on lung cancer using these patient groups as hospital referents will very likely result in underestimation of the true effect. On the other hand, certain patient groups may more likely avoid specific exposures, eg asthma patients avoiding environmental tobacco smoke (ETS) exposure, in which case using such patient groups as hospital referents may overestimate the effect of ETS on lung cancer. Using multiple referent groups with different disease diagnoses in different combinations may help to examine/minimise possible selection biases that could be introduced.

Information bias

Information on disease outcome is unlikely to be a problem, so long as the diagnostic criteria are objective and clearly defined. Ascertainment of outcomes would likely be similar in cases and referents, as the former are identified during routine medical care. It is possible that some referents, especially 'healthy' subjects in the community, can actually be harbouring the disease outcome, but have not yet been detected. This misclassification of some cases as referents will tend to bias any true associations towards the null, the extent of the bias will depend on how likely it is that true cases are present in the referent group. For rare (low prevalence) diseases (eg cancer), it should not be a concern.

Information on exposure(s) and potential confounding factors is obtained retrospectively, and

TABLE. Main sources of bias in case-referent studies*

Study design	Source of bias					Confounding
	Selection bias		Information bias			
	Investigator	Self (study subjects)	Exposure(s)	Outcome(s)	Confounding factor(s)	
Case-referent study: hospital-based						
Cases	+	+	++	+/-	+	++
Referents	++	+	+	+/-	+	
Case-referent study: population-based						
Cases	+	+	++	+/-	+	++
Referents	+	++	+	+	+	

* ++ denotes major source, + minor source, and +/- unlikely

if not already documented in records and abstracted using standard protocols by persons blinded to the disease status and hypothesis, such information can be subject to serious bias. In practice, a substantial amount of information on exposure(s) and potential confounding factors is obtained retrospectively through interviews and bias can originate from both the interviewers and the interviewees. Bias can be introduced by interviewers (interviewer bias), through unequal probing into or inaccurate classification of past exposures in cases and referents if they are not blinded to the disease status of the interviewees and the study hypothesis. In this situation, there may be an overestimation of associations between certain exposures and outcomes. Bias introduced by interviewees (interviewee or recall bias) can result in both overestimation and underestimation of associations. The former are not uncommon in studies examining hypotheses that are well known or prevailing in the population at the time, thus becoming a form of self-fulfilling prophecy. For example, health care workers suffering from severe acute respiratory syndrome in 2003 would be more likely to over-report touching their noses and eyes at work than those not infected. Underestimation of associations can occur in situations where the interviewees have no understanding of the hypotheses under study, but are asked to recall exposures (including potential confounding factors) in the past. Inaccuracies are bound to occur in recalling past exposures and information regarding the remote past is more likely to be inaccurate. Such inaccuracies (misclassifications) may result in both over- or under-reporting in both the cases and referents, but they are not more likely to occur in either group (non-differential). Hence, the resulting

bias will be towards the null hypothesis, ie no association.

Confounding

Confounding is especially of concern in examining risk factors for chronic non-communicable diseases (NCDs). This is because unlike infectious and genetic diseases, NCDs are usually multifactorial and not infrequently they have no single dominant contributing factors. As discussed in Workshop 2,² all known risk factors of an outcome can potentially lead to confounding when examining the association with a specific exposure. A well-conducted case-referent study takes into consideration all known risk factors for the disease under study and addresses their potential confounding effects. The first step in appraising whether confounding has been adequately managed entails assessing whether information related to all known risk factors (at least the major ones) has been collected. This is not an easy task unless the reader is very familiar with the subject area. Following this, it is necessary to examine whether known risk factors have been adequately adjusted for in the subsequent analyses and to check if there is enough statistical power for simultaneously adjusting for the number of known risk factors.

The Table summarises the major sources of bias in case-referent studies.

Ignatius TS Yu
Shelly LA Tse
 Clinical Epidemiology Group
Hong Kong Medical Journal

References

1. Yu IT, Tse SL. Clinical Epidemiology Workshop 3—Sources of bias in case series, patient cohorts, and randomised controlled trials. *Hong Kong Med J* 2011;17:478-9.
2. Yu IT, Tse SL. Clinical Epidemiology Workshop 2—General approach to critical appraisal of a medical journal paper. *Hong Kong Med J* 2011;17:405-6.