CLINICAL EPIDEMIOLOGY WORKSHOP

Workshop 3 — Sources of bias in case series, patient cohorts, and randomised controlled trials

By far the most common type of original paper published in the *Hong Kong Medical Journal* originates from a series of patients (cases) seen in a certain health care setting (such as a hospital and/or clinic). Not surprisingly, it is natural for research in clinical medicine to start with subjects seeking medical attention. The characteristics of a series of cases may just be simply described and summarised, eg socio-demographic backgrounds, clinical presentations, outcomes, etc, but the series of cases may also become the study subjects for conducting more elaborate research on prognosis and interventions. With a comparison or referent group, they can also be utilised for studies looking into risk factors or aetiology.

Case series

This study design refers to studies that mainly provide descriptive characteristics of the subjects (the cases). Collection of detailed information on clinical presentations and summarising the data systematically may provide support for identifying new diseases and guide clinical definitions, eg for the severe acute respiratory syndrome in 2003. Noting the socio-demographic characteristics of a series of cases, as well as the temporal and spatial distributions can sometimes provide a clue to risk factors and hence help generate a hypothesis. This can be tested subsequently with more elaborate analytic studies. In purely descriptive case series, confounding is not a concern, as the association between a certain factor and an outcome is not being studied. However, one should still look out for possible selection bias (ways the subjects were identified, selected, and included). If all necessary data are not routinely collected in the medical records and abstracted in a standard manner, information bias may also become a concern.

Selection bias

A case series consisting of all consecutive new cases seen in a hospital or clinic setting that fulfil predetermined inclusion and exclusion criteria should be representative of cases encountered at around the time the study is conducted. Since some diseases have seasonal variations, study periods spanning less than a year may not be sufficient and could mislead. Moreover, including all consecutive new cases seen in a hospital or clinic may not be representative of all cases in Hong Kong, but that is more a concern of external validity for generalisation of results. On the other hand, readers expect authors to provide information on the specific hospitals/clinics involved as well as the nature of the setting. By this means, they can make informed decisions as to whether the reported results are applicable to their own practice settings. Response and participation is usually not a problem for case series that purely describe clinical features (including laboratory findings), socio-demographic characteristics and treatments provided, as all such information is routinely collected for all attending patients and available for descriptive analyses. Nevertheless, if management of the disease involves invasive or risky investigations or interventions requiring specific consent (eg surgery), then it is expected that the participation rate (eg acceptance of surgery) among all eligible subjects be reported. If the outcomes during follow-up are also described, the follow-up rate should also be provided.

Information bias

Cases can be retrieved retrospectively or collected prospectively. Information on cases retrieved retrospectively is generally more objective, as it is collected routinely in relevant medical records. A standard form for abstracting relevant information by someone blinded to the hypothesis (if any) should provide unbiased information. However, missing or incomplete information could be an issue. For prospectively collected information, it is desirable to have standard protocols and/or forms to collect the necessary information, to avoid missing data for some patients. Furthermore, mechanisms should be built in to avoid excessive probing into particular aspects of a disease in certain patient subgroups, eg haemoptysis among chronic smokers.

Prognostic studies in patient cohorts

A case series can easily be extended to become a cohort of patients with the disease and then followed up to observe certain outcomes (prognosis), eg recovery, recurrence, death. By definition, a cohort study is an analytic study examining the association(s) between exposure(s) and outcome(s). Hence, simply describing the outcome(s) in a case series cannot be regarded as a cohort study. A cohort study looking at outcome(s) of patients can be called a prognostic study. As associations are being examined, confounding should be addressed in addition to selection bias and information bias.

Selection bias

The recruitment of patients into the cohort for a prognostic study is basically the same as for a case series, and hence selection bias on the part of the

investigator(s) is usually not a major concern, as long as clear inclusion and exclusion criteria are defined and all eligible cases are included. Response or participation rate among all eligible subjects should be reported in situations where consent for specific investigations and/or treatments is required, allowing readers to assess possible self-selection bias. Furthermore, as this type of study examines outcome(s) at time-points after the recruitment, loss to follow-up can be a major concern. If patients lost to follow-up are systematically different from those being successfully followed and if the proportion is substantial (eg >20%), the results can be seriously biased.

Information bias

In theory, a cohort study recruits and classifies subjects according to the presence or absence of a study factor, but for prognostic studies, subjects are recruited because they suffer from a certain disease. Factors affecting outcome(s) of interest (ie prognostic factors) do not usually form the basis for selecting patients for recruitment. Hence, information on these factors (including potential confounding factors when specific associations are being examined) would normally have been routinely collected in the usual medical care process and retrieved retrospectively (for historical cohort studies). Alternatively, in prospective cohort studies, relevant information could be specifically assessed and recorded using standardised protocols. Information bias on prognostic factors is usually not a major concern, provided information is objectively recorded, its retrieval is standardised, and the retriever is blinded to the hypothesis being tested. In historical (retrospective) studies, such blinding should also extend to the outcome status of subjects.

Information bias on the outcome is a major concern, unless it can be unequivocally objective, eg death from any cause. Hence, persons involved in assessing outcomes should be blinded to the hypothesis under study, and the methods or approaches used to ascertain the outcomes should be standardised for all patients. The latter is of particular importance, as the elapsed time to the event is frequently used as the outcome variable in analysing the associations with prognostic factors. If a certain subgroup of patients with a given prognostic factor is being followed up more frequently or receiving more investigations, it is inherently more likely that the outcome of interest (eg recurrence of a cancer) will be detected earlier, even though that particular prognostic factor does not influence the risk.

Confounding

Factors that independently affect outcome and at the same time may be associated with the prognostic factor being examined can result in confounding, and thus distort (bias) the association between the prognostic factor and the outcome of interest. In theory, any known prognostic factor can have the *potential* to confound associations between the outcome and another prognostic factor. Hence, such potential confounders should be considered and taken care of either in the recruitment process (using inclusion/exclusion criteria) or in the data analysis (by statistical adjustment). As a first step, information on known prognostic factors (especially those that are well documented and major) should be available and collected. Regrettably, such information is not always fully available in prognostic studies depending on historical cohorts. In theory, a prospective cohort is likely to be much better in this aspect, especially when it is well planned after a thorough review of the literature to identify known prognostic factors. At least three categories of prognostic factors should be examined for possible confounding: personal factors (eg age, smoking, co-morbidity), disease status (eg stage, biochemical or functional impairments), and treatment(s) received. Confounders in each of these categories commonly affect disease prognosis.

Randomised controlled trials

The gold standard to assess the efficacy or effectiveness of interventions is the randomised controlled trial (RCT). In the setting of clinical medicine, a RCT can be regarded as a special kind of cohort study in which patient subgroups are followed prospectively after having been randomly assigned to different interventions or exposures (including no intervention or placebo). Possible selection bias and self-selection bias can be present as in prospective prognostic studies; loss to follow-up being the most important cause of this problem. In RCTs without proper allocation concealment, there can also be serious selection bias. As the intervention or exposure is randomly assigned, information bias at this stage is unlikely to occur, although misclassification of the actual exposure can still ensue due to non-compliance with the assigned intervention. On the other hand, information bias with respect to outcomes is more liable to occur, especially if subjectively determined and if adequate blinding (of the patient, medical care provider, or outcome assessor) is not ensured. Theoretically, confounding is not an issue provided that the randomisation process is effective, as the comparison groups would be very similar with respect to prognostic factors other than the intervention.

A summary table is available in the online version on www.hkmj.org.

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

TABLE. Main sources of bias in studies on patient series*

Study design	Source of bias					
	Selection bias		Information bias			Confounding
	Investigator	Self (study subjects)	Exposure(s)	Outcome(s)	Confounding factor(s)	
Case series	+	+	+	+	N/A	N/A
Prognostic study	+	++	+	++	+	++
Randomised controlled trial	+	++	+/-	++	+/-	+/-

* N/A denotes not applicable, ++ major source, + minor source, and +/- unlikely