CLINICAL EPIDEMIOLOGY WORKSHOP

Workshop 6 — Sources of bias in cross-sectional studies; summary on sources of bias for different study designs

Cross-sectional studies

Cross-sectional studies are also called prevalence studies or surveys. These studies take snapshot views of the health status and/or behaviour of the study population at specified time-points. Examples include: the proportion of children being overweight and/or obese in different years, the proportion of adults having hypertension or diabetes mellitus at different ages in the year 2000, and the proportion of elderly (aged ≥65 years) having mild cognitive impairment or dementia at the time of the survey. Such surveys provide very useful information regarding the health status/behaviour of a population or a specific group, which is indispensable for assessing the health needs and planning for appropriate health services. Unfortunately, cross-sectional studies have sometimes been placed very low in the so-called 'hierarchy of evidence' and simply been neglected. However, this is the only type of study that can provide evidence on the health status of a specified population (group of **persons**) in a certain location (**place**) at a given time-point (**time**). A cross-sectional study can also be conducted to provide the baseline information for a prospective cohort follow-up to observe future new health outcomes. Cross-sectional studies are not meant for testing hypotheses and many people have expressed legitimate concern on the correct temporal relationships between alleged exposures and health outcomes discerned from these studies. High-quality cross-sectional studies are not easy to conduct, as they are susceptible to all three common sources of bias, of which selection and self-selection are of prime concern.

Selection bias

Unlike studies starting from a series of patients, there is often the need to select a sample of subjects from a pool (study population) that can be very large. Obtaining a representative sample is crucial but no easy task. Ideally, one should have the full list of eligible subjects and the sample is picked using simple random sampling with every subject standing an equal chance of being included. Multistage sampling and systematic sampling, with caution, can sometimes achieve the goal of having a representative sample, whilst being less dependent on the availability of the full list.[1] Unfortunately, samples are sometimes picked by convenience (being handy) and misleadingly reported as being a random sample. Such samples will not be representative and can result in serious selection bias. Stratified sampling is sometimes adopted to ensure adequate statistical power if subsequent stratified analyses or comparisons among strata are considered important. However, if the probability of being selected is different in the different strata (subgroups), weighting needs to be applied when reporting summary results for the whole study sample. Response rates from the sampled subjects vary, depending on many factors, but if not adequate (eg <80%) and the non-respondents are systematically different (from respondents), substantial self-selection bias can ensue. Cases with a certain health condition identified in cross-sectional studies are quite frequently compared to those without that condition using analyses similar to a typical case-referent (case control) study. Using such prevalent cases as opposed to incident cases can also introduce selection bias.[2]

Information bias

In cross-sectional studies, information on risk factors and health conditions (outcomes), as well as other factors, is often obtained at the same time-point. Adopting standardised and validated methods and using objective measures can help avoid information inaccuracies or biases. Self-reporting (answering a questionnaire or responding agreeing to be interviewed) is frequently used to collect a substantial proportion of the information in these studies. In which case, care must be taken in interpreting and drawing conclusions based on such information, as there is a tendency for respondents to provide what they believe to be socially acceptable answers rather than the truth, especially with regard to behavioural aspects and health conditions associated with taboos. Not infrequently, information on past exposures and behaviours (eg smoking history) is also collected, such that possible biases related to obtaining information retrospectively could be similar to those pertaining to case-referent studies.[2] In some studies, current exposures are examined for their effects on certain health outcomes, in which case such information can be regarded invalid, as the temporal relationship cannot be correctly clarified or no allowances may have been made for lag times. An example for unclear temporal relationship would

TABLE. Main sources of bias in major study designs*

Study design	Source of bias					
	Selection bias		Information bias			Confounding
	Investigator	Self (study subjects)	Exposure(s)	Outcome(s)	Confounding factor(s)	
Case series	+	+	+	+	NA	NA
Prognostic study	+	++	+	++	+	++
Randomised controlled trial	+	++	+/-	++	+/-	+/-
Case-referent study — hospital-based						
Cases	+	+	++	+/-	+	++
Referents	++	+	+	+/-	+	
Case-referent study — population-based						
Cases	+	+	++	+/-	+	++
Referents	+	++	+	+	+	
Historical cohort study						
Internal comparison	+	++	+	+	+	++
External comparison	+	+/-	+/-	+/-	+/-	++
Prospective cohort study						
Internal comparison	+	++	+/-	++	+/-	+
External comparison	+	+/-	+/-	+/-	+/-	++
Population cohort	++	++	+/-	++	+/-	++
Cross-sectional study	++	++	+	+	+	+

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

be whether current psychological stress measured by psychometric tools has led to or is a result of the reported musculoskeletal disorders. The lack of allowance for lag time is not uncommonly illustrated by using current dietary intakes (or nutritional status indicators) as possible risk factors for chronic diseases.

Confounding

A pure prevalence survey does not examine causal relationships, and hence confounding should not be a problem. However, associations between various factors and a certain health status are also frequently being examined. Alternatively, health indicators in various subgroups are compared, and under such circumstances, possible confounding by other factors known to be associated with the health outcome also needs to be attended to. Otherwise, the comparisons may not be valid.

Summary on sources of bias for different study designs

This is the last in the series of workshops discussing the sources of bias in studies with various study designs. The Table provides a summary that could alert authors, reviewers, and readers to the possible or probable sources of real or potential bias when dealing with papers reporting different types of studies.

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

References

1. Bland M. An introduction to medical statistics. 3rd ed. Oxford: Oxford University Press; 2002: 27-31.
2. Yu IT, Tse SL. Clinical Epidemiology Workshop 4—Sources of bias in case-referent studies. Hong Kong Med J 2012;18:46-7.