Evidence-Based Dentistry (2000) 2, 75–78

www.nature.com/ebd

toolbox

Causes, associations and evaluating evidence; can we trust what we read?

David R Moles[1] and Isabel dos Santos Silva[2]

[1]Clinical Lecturer and MRC Special Fellow in Health Services Research, Oral Pathology Unit, Eastman Dental Institute, University College London, London, UK; and [2]Senior Clinical Lecturer in Epidemiology, Cancer and Public Health Unit, Department of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, London, UK

As a profession we pride ourselves that our knowledge of oral conditions is derived from valid scientific inquiry. Similarly we like to believe that our treatments or interventions are based on good quality evidence of effectiveness. Secondary journals such as *Evidence Based Dentistry* attempt to seek out and review studies that are both informative and relevant to the day-to-day practice of dentistry. But just how valid is this 'evidence'? This article aims to briefly highlight a few important considerations when attempting to decide how much trust to place in any published work.

The starting point in this evaluation process is the fundamental concept that cause must be distinguished from the notion of *association*. Simply because two factors or conditions are associated, it does not necessarily follow that one is the cause of the other. For example, a study might show that patients who drank coffee were more likely to develop oral cancer than those who did not, but this does not necessarily mean that coffee intake is a cause of oral cancer. To assess whether an observed association is likely to be a true cause–effect relationship we need to consider the roles of bias, confounding and chance. We also need to assess whether it is consistent with what we know about causal mechanisms and biological processes. Each of these issues will be briefly discussed in the rest of this paper .

What is bias?
Bias is a systematic error. It leads to results that are consistently wrong in one or another direction. Bias leads to an incorrect estimate of the effect of a risk factor or *exposure* (eg sucrose consumption) on the development of a disease or *outcome* of interest (eg dental caries). The observed effect will be either above or below the true value.

1) Could the observed effect be due to bias?

Many types of bias have been identified, but the main types relate to how subjects were selected for inclusion in the study (*selection bias*) and how information on the relevant exposures and outcomes was measured or collected (*measurement bias*).

Selection bias occurs when there is a systematic difference between the characteristics of the subjects selected for a study and the characteristics of those who were not. For instance, selection bias will often occur with volunteers (self-selection bias). People who volunteer to participate in a study tend to be different from the general population. Thus, if the aim of a study is to estimate the prevalence of cigarette smoking in a particular town we should not rely on volunteers as they are more likely to be health-conscious and, hence, more likely to be non-smokers. Instead, we should randomly select a representative sample of the whole population. Similarly, it is important to consider why people might have withdrawn from the study before its completion. Is it because the treatment they were receiving was ineffective or uncomfortable in comparison with the alternative treatment? We will need to decide whether the results of the investigation were likely to have been compromised if one group of subjects had, on average, a shorter follow-up as a result of more people dropping-out.

The avoidance of selection bias is a major concern in the design of case–control studies (see box for details). In this type of study it is essential to ensure that controls are representative of the population from which the cases ori-ginated. Suppose a group of researchers is conducting a case–control study to assess the effect of cigarette smoking on oral cancer.

What is a randomised trial?
This is a study in which investigators intervene in the natural course of a disease or condition. Subjects are randomised to two or more study groups that differ only in terms of the intervention (eg preventive measure or treatment) under study. Subjects are followed over time to see if any outcome differences arise between the groups as a result of the different interventions they received.

What is an observational study?
For ethical reasons, randomised trials are limited to interventions that are believed to be of potential benefit. For instance, it would not be possible to conduct a trial to assess the effects of cigarette smoking on health (in which one group of people were asked to smoke). Researchers can *observe*, however, what happens among people who happen to be, or not be, smokers to see whether their disease risks are similar. This type of study is called observational study.

What is a cohort study?
This type of observational study is the one that most closely resembles intervention studies, except that allocation of subjects to the exposure is not determined by the investigator. The starting point in this type of study is the selection of a study population, or cohort. Information on the exposure status of each member of the cohort is collected at the start of the study and the entire cohort is then followed over time to assess whether the occurrence of disease in the exposed individuals is different from the occurrence in those not exposed. Cohort studies are often referred to as longitudinal studies.

What is a case-control study?
In this type of observational study, a group of subjects with the outcome of interest (called cases) is compared with a group without the outcome (the controls). For the subjects in each group, it is necessary to look back in time to establish whether they were exposed to the relevant exposure(s). Case-control studies are often referred to as being retrospective.

Cases were patients admitted to a certain hospital with a newly diagnosed oral cancer and controls were patients admitted to the same hospital with chronic bronchitis. A standard questionnaire was administered to both cases and controls that included questions on lifetime smoking habits. The researchers found no evidence from this study of an association between cigarette smoking and oral cancer. Can we accept their conclusions? The problem with this study is that the choice of controls was biased, as the prevalence of smoking among patients admitted with chronic bronchitis is likely to be much higher than among the general population resident in the catchment area of the hospital from which the cases originated. Consequently, the strength of the association between smoking and oral cancer was likely to have been underestimated in this study.

Randomised trials are less likely to be affected by selection bias as subjects are randomised to the groups to be compared. Randomisation eliminates selection bias on the part of the study participants and investigators, provided it is done after subjects have been determined to be eligible and have expressed willingness to participate in the trial. Methods based upon date of birth or date of entry have been used in some trials, with one intervention being assigned to those who were born (or who report) on even dates and another to those who were born (or who report) on odd dates. The problem with these methods is that it is possible for the investigators to know in advance the group to which a subject will be allocated. Therefore, conscious or unconscious bias may be introduced if this knowledge influences their decision on whether the subject is, or is not, eligible for entry into the study.

Measurement (*information*) bias occurs when the measurements of exposure and/or outcome are not valid (ie they do not measure correctly what they are supposed to measure). Errors in measurement may be introduced by the observer (*observer bias*), by the study individual (*responder bias*), or by the instruments (*instrument bias*) used to make the measurements (eg a badly-designed questionnaire). As a result of measurement error, study subjects will be

Examples of exposure and outcome misclassification
Exposure measurement is dependent on outcome status (exposure misclassification):
In a case-control study, an oral cancer patient may be more (or less) likely to report accurately their smoking habits than a healthy control patient. This would lead to an over- (or under-)estimation of the true effect of smoking on oral cancer. This type of bias can be minimised, to a certain exent, by keeping the study subjects 'blind' to the specific hypothesis being investigated.
Outcome measurement is dependent on exposure status (outcome missclassification):
A dentist may be more likely to diagnose pre-malignant lesions if they are aware that the subject is a heavy smoker. This would lead to an over-estimation of the true effect of smoking on oral cancer. This type of bias can be minimised by keeping the observer 'blind' to the exposure status of the study subjects.

misclassified in relation to their exposure and/or outcome status. This misclassification has particularly serious implications if the errors in exposure meausrement are related to the subjects' outcome status, or vice versa (see box with misclassification examples).

Bias is a consequence of defects in the design or execution of a study. Bias cannot be controlled during the statistical analysis of the data and cannot be eliminated by increasing the size of the study. The checklist highlights some of the key questions that help to identify potential sources of bias in published studies.

2) Could the observed effect be due to confounding?

Confounding is a term that describes the situation where an estimate of the association between an exposure and the disease is mixed up with the real effect of another exposure on the same disease, the two exposures being correlated. It is a difficult concept which might be illustrated with the help of the following example. Suppose we find that coffee drinkers have a higher risk of oral cancer than non-drinkers. Does it mean that coffee drinking causes oral cancer? The problem here is that there is an alternative explanation. Smoking is an independent risk factor for oral cancer and it is possible that people who drink coffee are more likely to smoke than those who do not. Perhaps the observed association is actually due to smoking habits, not coffee drinking (Figure 1).

Age and sex are the most common confounding variables in health-related studies; this is because these two variables are not only associated with most exposures we are interested in such as diet, smoking habits, physical exercise, etc., but they are also independent risk factors for most diseases.

Confounding can be dealt with at the design stage of an investigation by:

- *Randomisation* By randomly allocating subjects to study groups it is hoped that confounders are distributed equally between the groups. This is usually the most effective way of minimising the problem of confounding. If randomisation is properly done, it has the advantage that it controls for both known and unknown confounders provided the sample size is sufficiently large.
- *Restriction* This limits participation in a study to specific groups that are similar to each other with respect to the confounder (eg if smoking is likely to be a confounder then only non-smokers will be included in the study).
- *Matching* This selects comparison groups with similar backgrounds (eg non-smokers are matched with other non-smokers, while smokers are matched with other smokers). In practice, however, matching is only

Bias checklist

Selection bias
Was the study population clearly defined?
What were the inclusion and exclusion criteria?
Were refusals, losses to follow-up etc kept to a minimum?

In cohort and intervention studies
Are the groups similar except for the exposure/intervention under study?
Is the follow-up adequate? Is it similar for all groups?

In case-control studies
Did the controls represent the population from which the cases arose?
Was the identification and selection of cases and controls influenced by their exposure status?

Measurement bias
Were the exposures/outcomes of interest clearly defined?
Were the measurements as objective as possible?
Was the study 'blinded' as much as possible?
Were the observers or interviewers rigorously trained?
Were clearly-written protocols used to standardise procedures in data collection?
Were the study subjects randomised?
Was information provided by the patient validated against any existing records?

Adapted from dos Santos Silva, 1999[1]

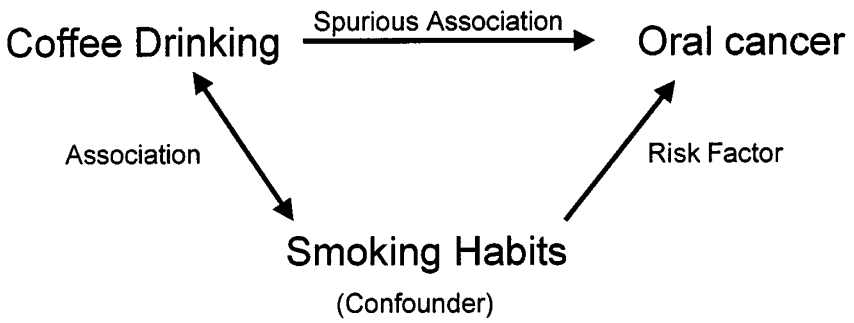

Figure 1 An example of confounding

done for well-known confounders such as age, sex, ethnic origin, place of residence and other socio-economic variables.

Confounding can also be controlled for in the *analysis* by:

- *Stratification* Here the strength of the association is measured separately in each well-defined sub-group (eg in the smokers and the non-smokers separately). The results are then pooled together using basic statistical techniques to obtain an overall summary measure of the association adjusted or controlled for the effects of the confounder.
- *Statistical modelling* These are more sophisticated mathematical techniques that simultaneously take into consideration the effects of all the possible confounders that have been recorded by the investigators.

It is only possible to control for confounders in the analysis if data on them were collected during the study. Obviously, the extent to which confounding can be controlled for will depend on the accuracy of these data.

3) Could the observed effect be due to chance?

The role of chance is assessed by performing statistical significance tests and, more importantly, by calculating confidence intervals. A proper discussion of these methods is beyond the scope of this article. It is, however, important to stress again that statistical methods cannot control for bias in the selection of subjects or in the measurement of the variables of interest.

4) Is the observed association consistent with causal mechanisms/processes?

'Causality' is almost impossible to prove as we can never be sure that the findings from a given study have not been affected by unknown sources of bias and confounding, or by chance. Thus, to help in our judgement of whether an observed effect is likely to be causal, the following aspects proposed by Hill[2] should be considered:

- *Time sequence* Did the exposure occur before the disease? Did the patient's ulcer occur as a result of some aspect of their diet, or did they change their diet because of the ulcer?
- *Plausibility* Is the association consistent with other biological knowledge? Are the findings consistent with animal experiments and our knowledge of the underlying biological mechanisms?
- *Consistency* Are the findings similar to those reported by other studies conducted in other settings and using different designs? The anti-caries effect of fluoride in drinking water was first established from observational studies that correlated the amount of naturally occurring fluoride in the water to caries levels in communities that drank the water. These results were later confirmed by intervention studies in which communities were artificially exposed to different levels of fluoride in the water.
- *Strength* A strong association is more likely to be causal than a weak one, eg if an exposure is associated with a 500% increase in risk as opposed to a 3% increase.
- *Dose–response relationship* Do people have a higher risk of disease if they had a higher exposure? The risk of developing oral malignancy increases with intensity of smoking as measured by amount smoked and duration of smoking.
- *Reversibility* Does removal of a possible cause result in reduced risk? The risk of lung cancer is reduced in ex-smokers when compared to current smokers.

These aspects should not be regarded as necessary conditions to establish causality. The only exception is time sequence – for an exposure to be a cause of a disease it clearly has to precede its biological onset. Because of the complexity of these issues it is rare that a single study will provide enough evidence that a certain exposure affects the risk of a particular condition. Usually, the findings need to be replicated by other studies with different designs and conducted in other settings.

Conclusion

Evaluating evidence from published studies is usually complex, with assessment of causality being just the starting point. Even after being convinced that the observed association is likely to be causal there are many other issues that should be taken into consideration before this finding can be translated into day-to-day practice. These include issues such as the extent to which the finding can be extrapolated to our particular population; whether the conditions and/or exposures investigated are common and/or serious enough to justify changes in practice; the cost-effectiveness of the intervention; the available resources and other competing priorities. Finally, it should be remembered that the publication of a paper in a peer-reviewed journal is no guarantee that the design, conduct, analysis, or conclusions of an investigation are correct. Although the information presented in this paper is not exhaustive, it is hoped that this brief review will be of some assistance to colleagues faced with the need to make evidence-based decisions everyday of their practising lives.

1. dos Santos Silva I. Cancer Epidemiology. Principles and Methods. Lyon: International Agency for Research on Cancer (IARC/WHO), 1999: 277–303.
2. Hill AB. The environment and disease: association or causation? Proc R Soc Med 1965; 58: 295–300.