EBN users' guide

Evaluation of studies of causation (aetiology)

CLINICAL SCENARIO

You have been appointed as the Nurse Manager of a nursing home that provides long term care for older people and have become aware that a large proportion of residents have pressure ulcers. You know that many preventive measures, such as special beds and mattresses, are promoted, but feel unsure as to which residents are most at risk and which characteristics predict those most likely to develop pressure ulcers. Although you know there are some pressure ulcer risk prediction tools available, you are not sure that these would apply to your residents and so would like to read some original research on the topic. Your long term goal is to ensure that care is targeted at those residents at highest risk so that preventable ulcers are avoided. Your focused clinical question is *which characteristics of nursing home residents place them at higher risk of pressure ulcer development?*

THE SEARCH

You begin by searching *Evidence-Based Nursing Online* (www.evidencebasednursing.com) using the search term "pressure ulcer*". This search identifies 8 abstracts, none of which address questions of risk factors. Similarly, a search of *Evidence-Based Medicine* finds no relevant studies. You try PubMed (www.ncbi.nlm.nih.gov), which is freely available online. You select the Clinical Queries search option, and the "Etiology" search because you are looking for articles concerned with the causation of pressure ulcers. You decide that your search should emphasise sensitivity over specificity so that you can minimise the risk of missing relevant articles. Your search terms are "pressure ulcer*" AND "nursing home*". This search identifies 32 abstracts, one of which, "A longitudinal study of risk factors associated with the formation of pressure ulcers in nursing homes" sounds relevant.¹

TYPES OF RESEARCH STUDIES

Studies that consider risk factors (often referred to as exposures) for certain diseases (often referred to as outcomes) are generally called analytical observational studies. They are distinct from randomised controlled trials because the researcher does not manipulate the exposure; rather, the exposure is merely measured and its association with the outcome is calculated. With respect to the clinical scenario, you are interested in the effect of patient characteristics (the exposures) on the development of pressure ulcers (the outcome).

The most common types of observational studies to assess risk factors for disease are cross sectional studies, case-control studies, and cohort studies. Each has distinctly different designs and differs in its advantages and disadvantages. Each of the study types will be described briefly in terms of a study that seeks to determine whether low body mass index (BMI) (exposure) is a risk factor for pressure ulcers (outcome).

In a *cross sectional study*, data on exposure and outcome are measured at the same point in time, on the same individuals. For example, data may be collected from a sample of residents from 5 nursing homes. Care providers would complete a questionnaire on each resident that would include information on weight and height (to calculate BMI), some

measure of the number and perhaps severity of pressure ulcers, as well as other factors that might be linked to pressure ulcers, such as age, recent hospital stay, chronic conditions, and mobility. These data would then be analysed to see if residents with low BMIs were more or less likely to also have at least one pressure ulcer.

In a *case-control study*, the researchers would identify a group of nursing home residents with pressure ulcers—the *cases*. They would also identify a group of nursing home residents who did not have pressure ulcers—the *controls*. The researchers would then collect information on previous exposures (eg, BMI on entry into the nursing home) for each case and control patient. The difference in the prevalence of the exposure (BMI) between the case and control patients would then be compared. Case-control studies look back in time to measure exposures and therefore are called *retrospective* studies. Case-control studies can also be "nested" within an existing cohort study.

In a *cohort study*, the researcher identifies a group of nursing home residents who do not have pressure ulcers and measures their BMIs. This group is then followed up over time to determine how many, and which, residents develop pressure ulcers.

The study we identified from our literature search is a prospective cohort study. Brandeis *et al* followed up a cohort of all new residents of 78 nursing homes over a 1 year period.¹ All residents included in the study did not have pressure ulcers at the time of nursing home admission or 3 months later when baseline measurements were done. This time delay was to ensure that risk factors external to the nursing home would not influence the study findings. All residents were then followed up for a further 3 months.

MEASURES OF EFFECT IN STUDIES OF CAUSATION

In studies of causation, we are interested in the relation between certain patient risk factors (exposures) and a particular condition or disease (the outcome). The relation between risk factors and outcomes is usually presented in terms of *relative risk*, that is, how much more (or less) risk do people with a particular characteristic have of developing the condition. Depending on the type of study and analysis, these relative risks are most commonly *risk ratios* or *odds ratios*. Let us return to the example of the relation between BMI and pressure ulcers. Hypothetical data are used in table 1 to show how results might be presented and the relative measures calculated.

We generally start by measuring the risk of having the outcome in the exposed group. In this case, we look at the first row of data in the table. The risk of having a pressure ulcer among those with low BMI is calculated by dividing the number of people who have pressure ulcers by the total number who have low BMI. Therefore, the risk of having the outcome in the exposed group is $116/1600 = 0.073$. We can multiply this number by 100, so that it becomes a percentage (7.3%). We then repeat this process for people without the exposure, those with high BMI (second row of data in the table). The risk of having the outcome among those without the exposure is $74/1400 = 0.053$ or 5.3%.

Table 1 Body mass index and pressure ulcers in a hypothetical sample of 3000 nursing home residents

		Pressure ulcer		
		Yes	No	Total
Body mass index	Low (<20kg/m ²)	116	1484	1600
	High (≥20kg/m ²)	74	1326	1400
	Total	190	2810	3000

In order to calculate the *risk ratio* we divide the risk of pressure ulcers among those with the exposure by the risk among those without the exposure. In this case, the risk ratio would be $0.073/0.053 = 1.37$. A risk ratio of 1.37 indicates that those with low BMI are at 1.37 times (or 37%) greater risk of having the outcome than those with high BMI.

In case-control studies, because we begin with people with and without the disease, we generally cannot measure disease incidence. In this instance, we use the *odds ratio* as the measure of the size of the effect of the exposure on the outcome. However, the calculation of odds ratios is not confined to case-control studies.

We calculate the odds in the exposed group by dividing the number of people with low BMI who have pressure ulcers by the number of people with low BMI who do not have pressure ulcers (first row of data in table). We then repeat this calculation for those without the exposure.

Odds in the exposed group = $116/1484 = 0.078$

Odds in the unexposed group = $74/1326 = 0.056$

We obtain the ratio by dividing the odds of having a pressure ulcer among exposed people (low BMI) by the odds of having a pressure ulcer among unexposed people (high BMI).

Odds ratio $0.078/0.056 = 1.39$

Risk ratios and odds ratios are interpreted in largely the same way; that is, those with the exposure have 1.39 times the odds of having the outcome. The odds ratio and risk ratio are similar when the frequency of the outcome is low, but they become increasingly divergent as the outcome becomes more frequent.

ARE THE RESULTS VALID?

In order to assess the validity of a study of risk factors (or causation) of a particular condition, we need to assess the extent to which 3 factors, namely chance, bias, and confounding, may have affected the study findings.

Magnitude of associations and the role of chance

In most studies of causation, we use statistical methods to assess the *role of chance*. We are interested in whether residents with a specific characteristic (eg, low BMI) are more likely to develop an outcome (eg, pressure ulcers) than residents without the characteristic. Statistical tests would test the null hypothesis that BMI is not a risk factor for pressure ulcer development in nursing home residents. The results of these statistical tests are commonly presented as p values and 95% confidence intervals (CIs).

By convention, the cut point for statistical significance for a p value is usually 0.05. This value indicates that it is only 5% likely that the observed relation is because of chance. Obviously, the smaller the p value, the less likely that an observed association between an exposure and outcome has occurred by chance. However, it is important to note that there is no clinical reason for assigning this level of significance.

Therefore, it is generally no longer seen as good practice to simply report a finding as “significant” or “not significant” or to indicate $p < 0.05$ for significant findings. Instead, actual p values and 95% CIs are preferred because they provide more detailed information about the strength of the evidence for rejecting or accepting the null hypothesis. It is important to note that p values do not provide information about the strength of the association between the exposure and outcome; this information is provided by risk ratios or odds ratios. p Values simply provide us with evidence to make a judgment about whether a finding may have occurred by chance alone. Statistical significance does not, therefore, relate to clinical importance. In some instances, we may interpret a non-significant difference between groups to be a clinically important difference based on our clinical experience.

Statistical significance is related to the *power* of a study, which is linked to sample size. If the sample size is not large enough to detect an effect of an exposure on an outcome, then there is a risk that we may calculate a false-negative result. We might conclude, based on this false-negative result, that an exposure (eg, BMI) is not related to an outcome (eg, pressure ulcer development), when it may simply be that the study was not large enough to demonstrate such a relation. Readers of study reports should look for a *sample size calculation* to indicate that the study population was large enough to minimise the likelihood of false-negative results.

Although cross sectional, case-control, and cohort studies allow us to assess the effects of several exposure variables at the same time, such *multiple hypothesis testing* can pose problems. That is, the greater the number of statistical tests we perform within a single study, the more likely we are to obtain false-positive results. We can minimise this risk by clearly stating, at the beginning of a study, which hypotheses we intend to test, with accompanying justification. Justification would generally come from previous research studies.

The study by Brandeis *et al*¹ followed from previous research that had examined risk factors for pressure ulcers, albeit using different study designs. Therefore, the risk factors that were considered in their study were justified on the basis of previous literature suggesting that these variables were clinically related to pressure ulcer formation. The authors generally presented 95% CIs around the estimates of effect but did not always present exact p values. However, the authors did correctly interpret the CIs as evidence to accept or reject the null hypothesis. Although the study had a large sample size ($n = 4232$), the authors did not report whether the study was sufficiently powered. This is an important consideration in this study because the sample was drawn from 78 nursing homes. Individuals living in the same nursing home are more likely to have similar characteristics and therefore cannot be assumed to be completely independent for the purposes of statistical analysis. A larger sample size would be required to investigate this type of “clustered” population.

Bias

Readers always need to know whether a study measured what it set out to measure; this is an issue of *internal validity*. In an epidemiological study, *bias* refers to any systematic error that results in an incorrect estimate of the association between an exposure and outcome.² Bias is a concept that is sometimes difficult to grasp because the ways in which it can occur and its effects on study findings can be difficult to interpret.

Let us return to our example of the relation between BMI and pressure ulcer development. In a cohort study of all nursing home patients in a given area, we collect self reported

Fig 1 Data as measured

data on patient weight and height and then follow up patients for 6 months to see if they develop ulcers. From our data on exposures (BMI), we are able to divide the cohort by exposure status into those with high (H) and low (L) BMI as represented in Figure 1. However, some of the nursing home residents reported their weight and height incorrectly and have been assigned to the wrong exposure category. The actual distribution of exposure status is represented in Figure 2.

The error in the information collected is an example of *information bias*. More specifically, this bias is referred to as *misclassification bias* because some patients have been misclassified as having a high BMI when in fact they have low BMI, and vice versa. The effect of this type of bias on the results of a study depends on whether the misclassification is dependent on the outcome. Our example used a cohort study design, which means that we did not know the outcomes of patients at the time the exposures were measured. Therefore, it is unlikely that knowledge of the outcome could have influenced the measurement of BMI. It is more likely that residents who reported incorrect BMI information were randomly distributed among those who developed pressure ulcers and those who did not. Therefore, this type of misclassification is known as *random* or *non-differential*. This has the effect of moving the size of the association between the exposure and the outcome towards the null—that is, the study is less likely to show a relation between BMI and risk of pressure ulcer development.

Let us consider a hypothetical case-control study of residents with and without pressure ulcers who are asked about previous exposures, including recent hospital stays. When asked to remember how many days they had been in hospital, some patients report this information slightly incorrectly. As before, this is a form of information bias. The fact that some patients had developed pressure ulcers made them think more about what had happened to them recently. Thus, residents with pressure ulcers were more likely to accurately report the length of hospital stays, whereas those without pressure ulcers tended to underestimate their length of stay. Because the misclassification of the exposure information is dependent on

Fig 2 True data

Table 2 Self reported length of previous hospital stay (exposure) and pressure ulcer development (outcome)

		Pressure ulcers	
		Cases	Controls
Long hospital stay (>10 d)	Yes	15	10
	No	5	10

Odds in the exposed group (long hospital stay = yes) = $15/10 = 1.5$
Odds in the unexposed group (long hospital stay = no) = $5/10 = 0.5$
Odds ratio = odds in exposed group/odds in unexposed group = 3.0

the outcome, this is known as *non-random* or *differential misclassification*. It is also known as *recall bias* because the misclassification is based on memory. Residents without pressure ulcers would be more likely to underestimate their length of hospital stay, and therefore, any association between increased length of stay and risk of developing pressure ulcers would be overestimated, as shown in the table 2.

We decide to check if the self reported number of days in hospital matched what was reported in hospital records. We do indeed find that residents with pressure ulcers (cases) accurately reported their length of stay, whereas those without pressure ulcers (controls) tended to underestimate their length of stay. This is an example of *differential misclassification*. Data from the hospital records and the resulting calculations are summarised in table 3.

Using the “true” data on hospital stay (from hospital records) showed that the original calculation *overestimated* the size of the association between longer hospital stays and pressure ulcer development. When the data from hospital records were used, the numbers in the exposed group (long hospital stay = yes) were more similar between the cases and controls than when self report data were used.

The other main category of bias is *selection bias*, which refers to errors in the process of identifying the study population. For example, in a case-control study, are exposed cases more (or less likely) to be selected than unexposed cases? In a cohort study, is allocation of exposure status related to development of the outcome? A further issue for cohort studies relates to the follow up of participants over time to see if they develop the outcome of interest. Readers need to consider whether there has been substantial loss to follow-up and, in particular, if those who dropped out differ according to the exposure, the outcome, or both. The length of the follow up period is related to the latency period of the outcome of interest. For chronic diseases, it may be necessary to follow up participants for several years before sufficient numbers develop the outcome. However, the longer the period of follow up, the more difficult it will be to ensure complete, or near complete, collection of outcome data.

Table 3 Length of previous hospital stay based on hospital records (exposure) and pressure ulcer development (outcome)

		Pressure ulcers	
		Cases	Controls
Long hospital stay (>10 days)	Yes	15	13
	No	5	7

Odds in the exposed group (long hospital stay = yes) = $15/13 = 1.15$
Odds in the unexposed group (long hospital stay = no) = $5/7 = 0.71$
Odds ratio = 1.62

Fig 3 Confounding

In effect, when we consider the role of bias, we are looking for alternative explanations for the study findings. Is there really a relation between BMI and pressure ulcers or could the observed finding be explained by some error in the measurement of BMI or pressure ulcers. Bias can be a major problem in observational studies, but can be minimised through good planning at the start of the study to ensure the sample is unbiased and use of the most objective outcome measures. These could include standardised instruments and questionnaires, validated for the relevant populations. This does not mean that objective measures are free of bias. A researcher who believes that BMI may be a cause of pressure ulcers may tend to round down the weights of participants known to have the outcome. More objective outcome measures can reduce opportunities for information bias. The risk of information bias can also be reduced by ensuring that outcome assessors are unaware of the exposure status of participants.

Unfortunately, once a study is biased, there is nothing the researcher can do about it. Thus, it is important to consider the potential role of bias at the planning stage of a study and to ensure that selection of participants and collection of data are done in ways that minimise bias.

In most cohort studies, the outcome status of participants is unknown at the time the exposure data are collected. For this reason, cohort studies are generally thought to be less prone to bias than case-control or cross sectional studies. Therefore, the findings of cohort studies are generally thought to be more credible than the findings of case-control or cross sectional studies. However, cohort studies may not be feasible to address all questions of causation. For example, a study of rare conditions would require an impractically large sample size to ensure that an adequate number of people would develop the outcome. Similarly, conditions with long latency periods would require impractically long follow up periods.

The study by Brandeis *et al*¹ was based on computerised data, routinely collected from all nursing home residents. These data were collected during assessments by trained nurses and were shown to be 90% reliable. The study used strict criteria for diagnosis of the outcome and did not include Stage I pressure ulcers in the analysis because of the potential difficulty of reliable identification. It was unclear who did the outcome assessments and whether they were blinded to participants' exposure status. No participants were lost to follow up, likely because they tended to remain living in the nursing home.

Confounding

Confounding is another form of alternative explanation for the findings of observational studies. A classic example of confounding is provided by studies that show a strong relation between coffee consumption and lung cancer. Does coffee drinking cause lung cancer? It is more likely that the relation between coffee drinking and lung cancer is confounded by smoking patterns: people who drink more coffee are also more likely to smoke, which is also related to lung cancer. In the assessment of confounding, we are interested in whether a third variable, which is associated with both the exposure and the outcome, could explain any observed

Table 4 Patient risk factors associated with pressure ulcer incidence

Risk factor	Odds ratio (95% CI)	p Value
Homes with a high incidence of pressure ulcers		
Ambulation difficulty	3.3 (2.0 to 5.3)	<0.001
Faecal incontinence	2.5 (1.6 to 4.0)	<0.001
Diabetes mellitus	1.7 (1.2 to 2.5)	<0.006
Difficulty feeding oneself	2.2 (1.5 to 3.3)	<0.001
Homes with a low incidence of pressure ulcers		
Male	1.9 (1.2 to 3.6)	<0.007
Ambulation difficulty	3.6 (1.7 to 7.4)	<0.001
Difficulty feeding oneself	3.5 (2.0 to 6.3)	<0.001

relation between these factors. Confounding is represented diagrammatically in Figure 3.

Unlike bias, if the effects of possible confounding factors are anticipated, they can be corrected during the analysis phase of a study. The simplest way to control for confounding is *restriction*. For example, in the example of coffee drinking and lung cancer, the effects of confounding could be controlled by restricting study participants to people who do not smoke. Unfortunately, many studies are affected by more than one confounding factor, which can make the application of restrictions difficult.

Case-control studies more commonly use *matching* as a technique to control for confounding factors. That is, controls are selected to ensure that the distribution of potential confounders is similar to that among the cases. Cases can be individually matched, for example, on age group and sex, or matched across the entire group of cases and controls. Although matching can improve the efficiency of a study, matching on too many factors is inadvisable as it may introduce selection bias, and of course, it would not be possible to determine the effects of matched variables on outcome. If matching is part of the study design, then this must be taken into account in the data analysis. For further information on matching, readers can refer to Rothman and Greenland.³

Stratification is another technique to examine the possible effects of a third variable on an outcome. Rather than restricting the sample (eg, excluding people who smoke), stratification allows researchers to examine separately the relation between coffee drinking and lung cancer in smokers and non-smokers, and then compare the 2 results. This provides us with an idea of the effect of coffee drinking on lung cancer, independent of smoking.

Several statistical techniques can be used to account for the possible effects of confounding factors. The benefit of such techniques, the most common of which are *regression* techniques (eg, logistic regression), is that several potential confounding factors can be considered simultaneously; we call this *adjusting* for each of the confounders. For example, Margolis *et al*⁶ were interested in whether men were at increased risk of developing pressure ulcers compared with women. The unadjusted rate ratio for this relation was 0.78, which means that men were 22% $([1.00 - 0.78] \times 100)$ less likely to develop pressure ulcers than women; the 95% CI of 0.70 to 0.88 (which does not cross 1) suggests that this finding is unlikely to be the result of chance. However, when potential confounding factors, including age and other medical conditions, were taken into account, the adjusted rate ratio for pressure ulcers in men compared with women was 1.01—that is, towards the null—with a 95% CI of 0.89 to 1.15. When age and medical conditions were taken into

account, the data no longer provided evidence of a relation between sex and pressure ulcers. These statistical techniques are also referred to as *multivariable* techniques because they consider the effects of several variables at the same time, each taking the others into account.

Brandeis *et al*¹ used a mixed approach to deal with confounding. Although the authors did not discuss the variable incidence of pressure ulcers across nursing homes in terms of confounding, they did, in a sense, attempt to account for these effects through stratification. The authors divided the data according to homes with high and low incidences of pressure ulcers and examined the variables associated with pressure ulcers for each group separately. However, they did not attempt to adjust for other potential confounding factors, such as number of chronic conditions or nutritional intake. The authors did mention in their discussion the possible effect of burden of disease as a way of explaining the observed sex differences in pressure ulcer risk. Although they did logistic regression analysis, they did not include these potential confounding factors in their regression model.

WHAT ARE THE FINDINGS?

Table 4 shows selected results from the study by Brandeis *et al*.¹ Several risk factors were identified as being related to the development of pressure sores in “high incidence” homes. For example, the odds ratio for the relation between ambulatory difficulty and pressure ulcers was >1.0 ; that is, residents with ambulatory difficulties had a greater risk of developing pressure ulcers than residents without ambulatory difficulties. In fact, they were 3 times as likely to develop pressure ulcers. The 95% CI of 2.0 to 5.3 around this odds ratio does not include 1 (the null), and therefore, we have evidence to reject the null hypothesis.

HOW CAN I APPLY THE FINDINGS TO PATIENT CARE?

The study by Brandeis *et al*¹ suggests several possible risk factors for the onset of pressure ulcers in the nursing home setting. 2 risk factors were found for both high and low incidence nursing homes—ambulation difficulty and difficulty feeding oneself. These 2 factors had relatively strong associations with pressure ulcer development: residents with either of these characteristics were at least 2.2 times more likely to develop pressure ulcers than those without these characteristics. Other risk factors were faecal incontinence and diabetes mellitus for high incidence nursing homes and male sex in low incidence homes. The study was a longitudinal cohort study, and therefore, we know that the outcome developed after the exposures and not the other way round (referred to as reverse causality). Because the outcome was not known at the time the exposures were measured, it is unlikely that there was any differential misclassification dependent on the outcome. Any misclassification of the exposures or the outcome is likely to have been random, which would have moved any observed associations toward the null. Nevertheless, the observed associations were quite strong. The study was perhaps weakest in terms of the way in which it dealt with confounding. For example, some of the observed associations may have been attenuated if the authors had controlled for the burden of disease among nursing home residents. However, it is unlikely that this would have explained all of the observed associations.

It is important to note that observational studies can only tell us about associations between exposures and outcomes. The findings do not necessarily inform us whether an

Summary of questions to assess the validity of studies of causation

Chance

- Were there clearly stated, justified, *a priori* hypotheses?
- Was there evidence to reject the null hypothesis (presentation of p values and 95% confidence intervals)?
- Was there evidence that the study was sufficiently powered?

Bias

- What was the study design? Cross sectional study, case-control study, or cohort study?
- In a case-control study, how was the sample selected? Is this likely to be associated the exposure?
- In a cohort study, how were participants allocated to exposure status? Is this likely to be associated with the outcome?
- Where possible, were objective or valid and reliable measurements used?
- In a case-control study, were the assessors of exposure blinded to outcome status?
- In a cohort study, were the outcome assessors blinded to exposure status?
- In a cohort study, was there substantial loss to follow up? What were the characteristics of those who left the study?

Confounding

- Did the authors consider possible confounding factors? Were they accurately measured?
- Did the authors analyse the data to take into consideration the effects of these potential confounders (restriction, stratification, or statistical analysis)?
- If matching was used, did the authors perform a matched analysis?

exposure actually *causes* an outcome. In order to be more confident of the causes of an outcome, we generally depend on summaries of evidence from several studies.

RESOLUTION OF THE CLINICAL SCENARIO

Based on the findings of Brandeis *et al*,¹ you suggest that residents with ambulatory difficulties be prioritised for the use of specialist beds or mattresses (an intervention that has been shown to be effective for the prevention of pressure ulcers⁵). You also decide to continue your research by identifying other studies (with various study designs) to see if they identify similar risk factors or perhaps suggest other higher risk groups.

JOY ADAMSON, PhD
Department of Health Sciences
University of York
York, UK

- 1 Brandeis GH, Ooi WL, Hossain M, *et al*. *J Am Geriatr Soc* 1994;**42**:388–93.
- 2 Unwin N, Carr S, Leeson J, *et al*. *An introductory study guide to public health epidemiology*. Buckingham: Open University Press, 1997.
- 3 Rothman KJ, Greenland S. *Modern epidemiology*. Second edition. Philadelphia: Lippincott-Raven, 1998.
- 4 Margolis DJ, Knauss J, Bilker W, *et al*. *Age Ageing* 2003;**32**:259–64.
- 5 Cullum N, Deeks J, Sheldon TA, *et al*. *Cochrane Database Syst Rev* 2000;(2):CD001735.