The clinical relevance of observational research

F E Lecky, P A Driscoll

In recent years there has been much discussion of evidence based medicine and of the importance of randomised controlled trials. It may appear that observational research is a lesser method of inquiry, belonging to darker times before the dawn of this new paradigm. In this review we aim to describe to the practising emergency physician the current relevance of observational research within an evidence based medicine framework. This will be done by defining the meaning of the term “observational study” and examining the nomenclature of observational methodologies (excepting qualitative research). The types of information provided by the different methodologies will be identified by using, where possible, examples from published studies in emergency medicine. There will be a particular emphasis on how observational studies can describe the performance of diagnostic tests. Finally we will discuss the limitations of observational research before concluding how this form of study can inform the clinical management of accident and emergency (A&E) patients.

Definitions and nomenclature

Observational study is best appreciated by considering its purpose and methods. In the broadest sense it aims to describe a health care situation relating to populations or groups of patients. In practice there is often a particular focus on the relation between two or more factors relating to a disease process. Observational studies should, like other methods of inquiry, collect data systematically and rigorously. However, unlike the randomised controlled trial and other experimental studies, no intervention is carried out by the researchers on the study subjects. Figure 1 shows the nomenclature differentiating the different types of observational study. The main division is between studies that describe a particular form of health care and those that use an observational approach to analyse the strength of association between factors. Studies therefore tend to be mainly descriptive or mainly analytical. These approaches are further subdivided according to whether the time frame is cross sectional or longitudinal, and in analytical studies, according to whether the study considers individuals or population groups.

Descriptive studies

Using a cross sectional approach, this type of study can determine the percentage of people in a population sample who have a characteristic of interest (for example, a disease) at a point in time. This percentage is known as the point prevalence. An example of this type of study was published recently in this journal. Johnstone et al identified whether or not children aged less than 13 admitted to hospital with a head injury had a skull fracture. Data were collected on children presenting to hospitals in the Edinburgh region between 1983 and 1989. It was found that 409 out of 3498 children had a skull fracture, giving a point prevalence of 11.7% at the time of admission to hospital.

In contrast, descriptive studies using a longitudinal approach often describe the incidence of a condition in a population sample. Incidence is a rate statistic which specifies the number of new cases of a disease occurring in a unit of population in unit time. This is commonly expressed as number of new cases per 100 000 head of population per year. The accuracy of the statistic will of course depend on completeness of follow up and quality and consistency of case ascertainment. An alternative longitudinal approach is one which repeats cross sectional samples from the same population over time to look at change in point prevalence. An example of this latter approach was again published recently in this journal. In 1995 Poznan- sky et al repeated the anonymous HIV testing of A&E attenders previously carried out in 1992 to 1993 and found that the point prevalence in the 16 to 45 year age group had risen from 1.3% to 3.3%.

Incidence is related to point prevalence (table 1), and therefore the increase in HIV prevalence described could be due to either an increased incidence of HIV infection or longer survival times.

Analytical studies

Analytical observational studies concentrate on population groups or individuals. Those focusing on population groups compare disease incidence rates, prevalences, or other morbidity statistics with the prevalence of possible associated factors between different groups or in the same group over time. An example of this is Wilkinson’s study of the relations...
Observational research

Descriptive surveys
- Longitudinal (studies of change)
- Cross sectional

Analytic surveys
- Group based (ecological studies, trend studies)
- Individual based
  - Longitudinal (time span)
  - Prospective (cohort, forward looking)
  - Retrospective (backward looking)
  - Cross sectional (non-directional)

Figure 1  Nomenclature of types of observational study (from ref 8).

between income distribution and life expectancy in Western countries. The percentage of national income received by the poorest 70% of the population was found to be related to average life expectancy (fig 2). From this it was hypothesised that relative deprivation contributed to population morbidity and mortality.

Analytical studies that focus on individuals also look at disease and possible associated factors. However, they use specific information from subjects in the sample and may attempt to imply cause and effect. This type of observational study can employ different time scales and therefore different methods for selecting patients.

Retrospective or cross sectional studies tend to select subjects on the basis of whether they have the outcome of interest or not. These are therefore known as “cases” or “matched controls” (hence the term case–control study). It is then determined which of these subjects have had the exposure of interest. Prospective or cohort studies sample from a prespecified population and follow the subjects up to determine how many of those with and without the exposure go on to develop the outcome of interest (fig 3).

An example of a retrospective or case–control study was one which assessed the effect of sleeping position on the risk of sudden infant death syndrome (SIDS). As such rare important conditions would require enormous resources to study prospectively, this study used cases matched to controls by age and date of interview. The sleeping positions of cases and controls were then compared (table 2).

From this type of data it is possible to calculate the odds ratio (OR) for SIDS with side or front sleeping as opposed to sleeping on the back:

\[
OR = \frac{\text{odds of SIDS with side or front sleeping}}{\text{odds of SIDS with sleeping on the back}} = \frac{106/265}{85/509} = 2.48
\]

This is not quite the same statistic as the risk ratio which gives the ratio of disease incidence between exposed and unexposed populations. To obtain incidence rates, a study needs to be prospective and therefore more resource intensive in design. Consequently retrospective studies are more suited to describing the effects of potential hazards on the incidence of more common diseases such as cancer or ischaemic heart disease. One of the most famous examples of this form of study is the work of Doll and Peto on the relation between smoking and the incidence of death from lung cancer in British doctors.

Table 2  Case–control example: infant sleeping position and risk of sudden infant death syndrome (SIDS) (from reference 3)

<table>
<thead>
<tr>
<th>Sleeping position</th>
<th>SIDS</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Side or front</td>
<td>106</td>
<td>265</td>
</tr>
<tr>
<td>Back</td>
<td>82</td>
<td>509</td>
</tr>
</tbody>
</table>

Figure 3  Analytical observational studies relating exposure to outcome.

The relative risk (RR) of dying from lung cancer in a year for those smoking ≥ 25 cigarettes a day compared to non-smokers can thus be calculated:

\[
RR = \frac{\text{incidence of death from lung cancer in smokers } \geq 25}{\text{incidence of death from lung cancer in non-smokers}} = \frac{208/7}{52.7} = 3.49
\]

For the same example a 2 × 2 table for the calculation of an odds ratio can easily be reconstructed:

Excel table showing the odds ratio calculation with the following data:

<table>
<thead>
<tr>
<th></th>
<th>SIDS</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Side or front</td>
<td>106</td>
<td>265</td>
</tr>
<tr>
<td>Back</td>
<td>82</td>
<td>509</td>
</tr>
</tbody>
</table>

The odds ratio (OR) is calculated as:

\[
OR = \frac{106}{265} \div \frac{82}{509} = 2.48
\]
OR = odds of death from lung cancer in smokers ≥ 25 : odds of death from lung cancer in non-smokers
= (208/99 792) : (7/99 993)
= 29.71

In this example the odds ratio is slightly greater than the relative risk but the increase is so small as to have no clinical significance. However, as outcomes (death from lung cancer in the example) become more frequent in the population, the degree to which an odds ratio overestimates the relative risk from an exposure increase, and becomes more significant. This phenomenon should also be borne in mind when critically appraising experimental studies because trials of treatment may use odds ratios as a measure of clinical effectiveness.

Studies evaluating diagnostic test performance
These studies are a type of observational analytical study usually performed on a cohort of patients with a common presentation such as chest pain. Patients are followed up from the time of presentation. Subsequently, the initial diagnostic test result is compared with a later gold standard marker of disease, which often does not become available until later in the patient’s clinical course. Sensitivity and specificity of the initial test for the disease are commonly used as indicators of performance. An example of this type of study is given in a paper by Sackett et al. The investigators examined creatinine kinase (CK) as a diagnostic test for myocardial infarction in chest pain patients admitted to hospital, where ECG, angiography, and necropsy findings were the later gold standard of diagnosis. The results are shown in table 4, where a CK value of ≥ 80 IU indicates a “positive” CK test.

The sensitivity of a diagnostic test is the proportion (percentage) of patients with the target disorder who have a positive test result. This is equivalent to 215/231 or 93% from table 4. The specificity of a diagnostic test is the proportion (percentage) of patients without the target disorder who have a negative test result. This is equivalent to 114/130 or 89% from table 4.

The positive predictive value from a diagnostic test result is the proportion of patients with positive results who have the target disorder. From table 4 this is equivalent to 215/231 or 92%. The negative predictive value from a diagnostic test result is the proportion of patients with negative results who do not have the target disorder; again from table 4 this is equivalent to 114/129 or 89%.

Finally the accuracy of a diagnostic test is the proportion of all test results that are “true positive” and “true negative,” which from table 4 is equivalent to (215 + 114)/360 or 92%.

It is noticeable from table 4 that a clinician, before performing CK enzyme tests on a patient (from this population), could be aware that on average two thirds (230/360) of this type of patient would later prove to have a myocardial infarct. Therefore the patient’s pre-test probability of myocardial infarction is ~67%. If this is the case then a CK of ≥ 80 IU with a positive predictive value of 92% gives a 25% post-test increase in the probability of myocardial infarction. However a negative CK gives an 11% probability of myocardial infarction (15/129), a 56% reduction from the pretest probability. Therefore a negative result in this study would almost rule out a myocardial infarct. Conversely a positive result sufficiently increases the probability of myocardial infarction for the patient to be treated as having this diagnosis.

Clinicians can often estimate from the history and examination the pretest probability of disease. This is important as it can influence the later “post-test” probability of disease—that is, the positive and negative predictive value. This is illustrated in table 5, which shows hypothetical data. The lower prevalence of myocardial infarction (20% in table 5) is perhaps more reflective of that in A&E patients presenting with chest pain.

Sensitivity and specificity are generally considered to be constant features of diagnostic test performance and hence, at 93% and 89% respectively, they determine the numbers given in bold in table 5. However, the predictive values, calculated as described above, have altered. The post-test probability (positive predictive value) for myocardial infarction with CK ≥ 80 IU is 69% (19/27), a figure markedly reduced from the previous example. The negative predictive value of CK < 80 also alters, being increased to 98% (72/73).

For this reason the probability of disease from diagnostic test results has to be interpreted in the context of pretest information. Though intuitively this sounds sensible, it is easy to be seduced by “absolutism” from laboratory or radiological investigation. Well conducted studies have shown that clinicians ignore important information from the history and examination at their peril. This bedside information is often more powerful than laboratory tests for altering the probability of disease to a level at which the patient can

<table>
<thead>
<tr>
<th>Number with myocardial infarction</th>
<th>Number without myocardial infarction</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number with CK ≥ 80 IU</td>
<td>19</td>
<td>8</td>
</tr>
<tr>
<td>Number with CK &lt; 80 IU</td>
<td>1</td>
<td>72</td>
</tr>
<tr>
<td>Total</td>
<td>20</td>
<td>80</td>
</tr>
</tbody>
</table>
reasonably be treated or safely discharged. This philosophy of interpreting the likely “truth” from new information, in the context of what is already available, was first advocated by Bayes and forms the basis of Bayesian statistics. Hence diagnostic test results are not absolute predictors of disease by a given probability but have likelihoods which can be combined with pretest information in a relatively simple mathematical way. This is achieved through the concept of diagnostic test results that have a likelihood ratio as an indicator of performance.

Likelihood ratios
The derivation of likelihood ratios is relatively straightforward. From table 5, the likelihood ratio for myocardial infarction where CK > 80 IU (likelihood ratio positive, LR+) = the ratio of the likelihood that someone with a positive result has the disease / the likelihood that someone with a positive result does not have the disease:

\[
\text{LR}^+ = \frac{19/20}{8/80} = 0.5
\]

The likelihood ratio for a “negative” CK (< 80 IU, LR−) is calculated in a similar way:

\[
\text{LR}^- = \frac{1/20}{72/73} = 0.05
\]

Recalculation of the likelihood ratios using data from table 4 gives the same results.

Consequently, like sensitivity and specificity, the likelihood ratios are non-prevalence-dependent indicators of diagnostic test performance. More importantly they can also be combined with prevalence or pretest probability, using easy maths to give post-test probability. The principle is given by the following equation, shown applied to the clinical situations with positive test results from tables 4 and 5:

Pre-test odds of target disorder × likelihood ratio from test result = post-test odds of target disorder

Table 4: 2.0 × 9.5 = 19.0
Table 5: 0.25 × 9.5 = 2.38

Odds are relatively simple to convert into probabilities: if one thinks of the example that a probability of 80% (0.8) is equivalent to 4:1 odds, then it is clear that odds = probability / (1 − probability), and probability = odds / (odds + 1).

Thus it is shown that likelihood ratios are robust indicators of diagnostic test performance which can be used for the different clinical presentations encountered by the emergency physician.

Limitations of observational research
The major weakness of observational research is the possibility that bias and confounding will influence the results of studies. In theory this should be less of a problem for experimental research. Bias means that the results of a study are different from “the truth” because of the way in which the study has been conducted.

There are two major types of bias. Selection bias means that some or all of the study subjects were selected in a way which leads to overrepresentation of certain groups. For example, in the study of SIDS, cases and controls were not matched for social class. It is possible, then, that the study result could have been affected by the different rates of SIDS between different income groups. This type of observational study (case–control) may also be prone to the other major source of bias, information bias.

Information bias is caused by shortcomings in collecting and recording data. An example of this is recall bias concerning the exposure of interest. This is because it is recognised that “cases” with a disease may be more likely to recall that they have had a given exposure (to an environmental hazard) than controls. The different varieties of selection and information bias are detailed in reference texts.

Confounding describes an occurrence whereby an association found between two factors may be incidental rather than cause and effect. A causal association is made because of the failure of the study to measure and take account of other factors which can be associated with both the purported cause and effect. An example of this phenomenon is found in studies which have looked at the relations between mortality and the grade of doctor managing the patient. This work shows that death is more likely if you are treated by a consultant than a junior house officer (UK Trauma Network, personal communication). If this crude unadjusted information is taken at face value bizarre inferences can be made. These could include the deduction that the training of junior doctors is counterproductive. However, this association is confounded by the severity of illness factor which is associated both with increased mortality (outcome) and a junior doctor (the exposure) being involved in the patient’s care (fig 4). A proper analysis of these data would therefore have to compare the outcomes of
patients with the same severity of illness treated by junior doctors and consultants.

Given these potential weaknesses, how can the strength of evidence from observational research be weighted against other forms of inquiry? There is no doubt that for studying the effects of treatment, observational methods come second best to the randomised controlled trial. However, good observational research design can avoid bias and ensure that known confounders are adjusted for in the analysis. Some reference texts give check lists to aid the reader in determining the reliability of different types of study (table 6).

**Summary**

In this review we have described some uses of observational research, and some of its limitations. The examples given have shown that observational studies can be an ethical and practical method of inquiry. The major uses of observational research include determining the incidence and prevalence of disease, aspects of disease aetiology, the effect of exposure to known hazards, and the application of diagnostic tests in the clinical setting. Observational studies are most useful to the emergency physician for determining the performance of diagnostic strategies. As this is a specialty where clinicians have just a single diagnostic opportunity, high quality observational studies must continue to help develop evidence based emergency medicine.

We gratefully acknowledge the permission granted by D L Sackett (tables 4 and 6), R G Wilkinson (fig 2), and J H Abramson (fig 1) for the reproduction of their work, and that of their respective publishers: Little Brown and Co, BMJ Publishing Group, and Churchill Livingston.

**Further reading**


---

**Faculty of Accident and Emergency Medicine**

Consultant appointments, September to December 1997

<table>
<thead>
<tr>
<th>Consultant</th>
<th>Hospital</th>
</tr>
</thead>
<tbody>
<tr>
<td>Miss B Brooks</td>
<td>Royal Surrey County Hospital, Guildford</td>
</tr>
<tr>
<td>Mr Paul Cornelius</td>
<td>Conquest Hospital, St Leonards on Sea</td>
</tr>
<tr>
<td>Dr K Henderson</td>
<td>Homerton Hospital</td>
</tr>
<tr>
<td>Dr Stephen Meek</td>
<td>Royal United Hospital, Bath</td>
</tr>
<tr>
<td>Dr M Mitchell</td>
<td>Hillingdon Hospital, Uxbridge</td>
</tr>
<tr>
<td>Mr Iomhar O'Sullivan</td>
<td>United Bristol Healthcare Trust</td>
</tr>
<tr>
<td>Mr V Reece</td>
<td>South Tyneside District Hospital</td>
</tr>
<tr>
<td>Mr M Saab</td>
<td>Fairfield General Hospital</td>
</tr>
<tr>
<td>Mr J D Whitaker</td>
<td>Sharoe Green Hospital</td>
</tr>
<tr>
<td>Mr J H Williams</td>
<td>Carmarthen and District NHS Trust</td>
</tr>
</tbody>
</table>

---