Epidemiology for ophthalmologists: an introduction to concepts, study designs, and interpreting findings

H Kuper, C Gilbert

This review provides an overview of the types of information epidemiological research can provide and how these data can be used. The aim is to provide the readers with basic epidemiological skills to allow them to read critically scientific articles and to gain proficiency in communicating about epidemiological research. All examples in the review are drawn from the ophthalmic literature. The first part of the review is relatively conceptual and focuses on epidemiological theory, including case definition, measures of the burden of disease, sampling and the interpretation of results. In the second part different study designs are described—specifically, cross sectional surveys, cohort studies, case-control studies, and randomised controlled trials, and the strengths and limitations of each highlighted.

The intention of this paper is to give an overview of the types of information epidemiological research can provide, and how these data can be used. Broadly, epidemiological research entails assessing the burden and aetiology of disease, and the efficacy of preventive measures or treatments in populations. This information has a variety of uses (table 1). Epidemiological research often relates closely to other areas of research so that findings generated in one sphere are explored, tested, or evaluated in another (fig 1). Epidemiology has been placed unashamedly at the centre of this interconnecting web of endeavour, as the information has such a key part to play in public health and service provision—or should do.

The first part of this review focuses on epidemiological research issues, measures of the burden of disease, and causation. In the second part we discuss different study designs, highlighting the strengths and limitations of each, using examples from the ophthalmic literature.

EPIDEMIOLOGICAL CONCEPTS

Epidemiological studies address questions such as: “How common is the disease?” “Who is most affected?” “What is the underlying cause?” and “What can be done to prevent or treat it?” These questions are answered by conducting studies on a sample of people. The purpose of any study must be clearly articulated, as all aspects of the design, methodology, and analysis flow from this (fig 2).

Case definition

Any epidemiological investigation starts with the definition of the disease or condition of interest, the question of “When is a case a case?” In terms of blindness, a person is defined as blind by the World Health Organization if they have a presenting visual acuity (VA) of <3/60 in the better eye, or a central visual field of <10°. If the VA is <6/60–3/60 the person has severe visual impairment (SVI), and if <6/18–6/60 they are visually impaired (VI). With these definitions, globally in the year 2000 there were about 50 million blind people, 25 million with SVI and 125 million with VI. Changing the definition of blindness will change the estimated burden of disease. If blindness were VA <6/60 there would be 75 million blind people globally, rather than 50 million. We must, therefore, be clear about what constitutes a case and what does not.

Populations and sampling

Measuring the condition of interest in an entire population is not usually feasible. For ease and speed the condition of interest is measured only in a sample taken from the target population, and the findings are extrapolated back to this population. Population can mean the whole population of a country or region, or (more often) the population defined by particular characteristics (for example, diabetic women).

Measures of the burden of disease

The most commonly used measures of disease burden are prevalence and incidence. Prevalence is the proportion of the population with the condition of interest. (A common mistake is to use the term “prevalence rate” rather than “prevalence.” Since prevalence is a proportion, not a rate, the term “prevalence rate” is incorrect and should be avoided.)

Prevalence estimates are used to calculate the expected number of people with disease in the population. In a recent national survey of blindness in Bangladesh, a sample of 11 624 individuals aged ≥30 were examined, of whom 162 had a VA of <3/60 in the better eye. The prevalence of blindness was therefore 162/11 624, or 1.39%, giving a national (age standardised) estimate of 650 000 blind people in those aged ≥30.

Incidence differs from prevalence as it relates to new, rather than existing, cases. Broadly, there

Abbreviations: ARMD, age related macular degeneration; CR, cumulative incidence ratio; IRR, incidence rate ratio; RCT, randomised controlled trials; SVI, severe visual impairment; VA, visual acuity; VI, visual impairment
are two measures of incidence: cumulative incidence and incidence rate. Cumulative incidence is the number of new cases that occur in a population which is disease free at baseline (that is, excluding prevalent cases) over a specified period of time. Approximately seven million people become blind every year—that is, there are seven million incident cases of blindness per year. The disease free population at risk is 6000 million (global population) minus the number of existing cases of blindness (50 million), giving 5950 million. The global cumulative incidence of blindness in the year 2000 was therefore approximately seven million divided by 5950 million, or 0.1%.

The incidence rate uses person time at risk as the denominator, rather than the population at risk. During the follow up of a group of people who are disease free at baseline, each person may (a) develop the disease of interest, (b) be lost to follow up, (c) develop a competing disease or die so that they are no longer at risk of becoming a case, or (d) remain disease free. To calculate person time the clock starts ticking at enrolment and stops when one of these events occurs or follow up ends. The total person time at risk is the sum of the person time at risk for each subject in the study. Ten person years of follow up can be produced by one person followed for 10 years, 10 people each followed for 1 year, or 20 people each followed for 6 months. The person time of follow up therefore depends on the number of subjects and the amount of time they are each followed. To calculate the incidence rate we divide the number of incident cases that occur during follow up by the total person time at risk. Incidence rate is expressed as cases per unit of person time (for example, four per 100 000 person years).

Table 1

<table>
<thead>
<tr>
<th>Study design</th>
<th>Type of information collected</th>
<th>How this information can be used</th>
</tr>
</thead>
</table>
| Descriptive and analytical studies | ● Burden of disease (prevalence and incidence)  
 ● Distribution of disease, with  
 ● Identification of high risk groups  
 ● Risk factors for disease (aetiology) | Policy  
 Advocacy  
 Priority setting  
 Planning and resource allocation  
 Target setting  
 Evaluation  
 Impact assessment |
| Intervention studies          | ● Effectiveness of preventive measures  
 ● Effectiveness of treatments | Health promotion  
 Preventive measures  
 Therapeutic interventions |
| Meta-analysis and systematic reviews | ● Summary of evidence of effectiveness of preventive measures  
 ● Summary of evidence of effectiveness of treatments | Policy  
 Guidelines  
 Planning |
Relative risk refers to either CIR or IRR. Another relative measure is the odds ratio, where the odds of exposure are compared between those with and those without disease to estimate the exposure disease association. For relative measures, a value of 1.0 indicates no association between exposure and disease. A value >1.0 shows that there may be a positive association between exposure and disease (exposure may cause disease), and a value <1.0 indicates an inverse association between exposure and disease (exposure may protect from disease).

**Chance and sample size**

An estimate measured in a sample should reflect the level of certainty that we have in this estimate. We can measure the prevalence of myopia in a sample of 20 children randomly selected from a population of 200 children. However, if we randomly select a new sample of 20 children the estimated prevalence may differ. This sampling variability is the result of chance and is an inevitable consequence of measuring disease in a selected sample rather than in the whole target population. We express the degree of confidence that we have in our estimate by using confidence intervals which provide an upper and lower range for the estimate. In the national survey of blindness in Bangladesh the 95% confidence interval for the estimate of the number of blind ranged from 524,000 to 725,800 blind people. The interpretation of 95% confidence intervals for the sample estimate is that the “true” value for the population lies within this range 95% of the time. Confidence intervals depend on the size of the study sample: as the sample size increases we are more confident that the findings in our sample reflect the “truth” and so the estimates will have narrower confidence intervals. If the study is too small, the 95% confidence intervals of the estimates will be wide, giving imprecise measures of disease burden, or failing to show a difference between two groups even if one does exist. Although they are distinct concepts, when the 95% confidence interval does not include the null value this usually reflects a statistically significant result.

**INTERPRETING RESULTS**

We do not conduct studies to make inferences about a specific sample, but to find out general “truths” about whether an exposure is associated with a disease, or an intervention prevents a disease or delays its progression. To begin to make general inferences from the results of individual studies we must be confident that the results were not the result of chance, bias, or confounding (fig 3).

**Chance**

Variations due to chance are an inevitable consequence of sampling, but the effects can be minimised by having a study that is sufficiently large. The term “power” is often used and this refers to the likelihood of the study resulting in a statistically significant finding, given that there really is a “true” difference between our comparison groups. If a study does not have a statistically significant result one is left wondering whether this is because the sample was not big enough to detect it or there really is no difference.

**Confounding**

All associations are potentially influenced by the effects of confounding, which can be thought of as alternative explanations for an association. We can best describe this in stages.

(a) An exposure appears to be associated with a disease.

(b) But the exposure is associated with a third factor, the confounder, which is an independent cause of the disease (a variable on the pathway from exposure to disease is *not* a confounder).
have been confounded by poverty. The association between sunlight and cataract may, therefore, independently be related to the prevalence of cataract. The more likely to be exposed to sunlight, and poverty is positively correlated and argued for the avoidance of sunlight confounding can lead to the finding in the sample (result of the study) being different from the “truth” in the target population.

A study was undertaken in Australia to assess whether ultraviolet radiation (exposure) was associated with cataract (disease). The authors reported a statistically significant positive correlation and argued for the avoidance of sunlight as a preventive measure. However, less wealthy people were more likely to be exposed to sunlight, and poverty is independently related to the prevalence of cataract. The association between sunlight and cataract may, therefore, have been confounded by poverty.

If everyone in the study had the same status with respect to the confounder (for example, all came from the same socioeconomic group) then the factor could no longer exert a confounding influence. This principle is used to control for confounding. In the design, confounding can be controlled by restricting the sample to people with certain characteristics (for example, the same socioeconomic status) or through randomisation (discussed later). In the analysis, multivariate regression or stratified analyses are used as a means of keeping confounders constant, and produce estimates that are “adjusted” for confounder(s). For instance, in an age adjusted analysis we control for age to reduce its confounding effect on the exposure disease association.

Factors which should be considered in interpreting (and hence also in designing) epidemiological studies:

**Chance**
- Can be measured by confidence intervals
- Cannot be prevented
- Can be reduced by an adequate sample size

**Confounding**
- Can be reduced by good study design, and taken account of during data analysis

**Bias**
- Effect on findings cannot be measured
- Can be reduced by good study design (selection bias) and quality control during study (information bias)
- Not influenced by data analysis

The “truth” in the target position

Extrapolation from estimate in sample to target population

Estimate in the sample (ie, result of the study)

Figure 3 Schematic representation of how chance, bias, and confounding can lead to the finding in the sample (result of the study) being different from the “truth” in the target population.

Bias
Study findings may be influenced by bias, which is the deviation of results from the truth. There are two main types of bias: selection bias and information bias (also called measurement bias). Selection bias is the error caused by systematic differences in characteristics between those who take part in a study and those who do not.

Causation
If the effects of confounding, chance, and bias are ruled out then causation becomes a possible explanation for the exposure disease association. Bradford-Hill outlined guidelines for causality, arguing that while we cannot prove causality, there is a stronger case if certain criteria are fulfilled.

Causation
The case for causation is strengthened if the association between the exposure and disease is large (strength of evidence), consistent across different studies (coherence), in line with current biological knowledge (experimental evidence) and generally credible (plausibility). Showing that increasing the level of the exposure further increases the risk of disease also supports the case for causality (dose-response relation). Bradford-Hill also argued that a role for causation was strengthened if a specific exposure leads to one specific disease (specificity), but since some exposures (for example, smoking) can lead to many diseases this criterion is weak and will not be discussed here.

STUDY DESIGNS
The study designs used in epidemiology can be grouped as descriptive or analytical studies (cross sectional, cohort, and case-control studies), where investigators observe the events as they unfold naturally, and intervention studies (randomised controlled trial), where investigators manipulate exposure(s) to risk factors/treatments to assess their impact on disease.

Cross sectional survey
Cross sectional surveys allow us to measure the prevalence of disease, so that the burden of disease can be estimated. To conduct a cross sectional survey, we carefully sample the desired number of study participants and examine and/or interview them to determine whether they have the disease and exposure(s) of interest. These data can be used to calculate the overall prevalence or to compare the prevalence in the exposed and unexposed groups and thereby explore aetiological questions.
Cohort studies allow us to measure predictors of disease incidence. To conduct a cohort study, a group of people free from the disease of interest are recruited and characterised as "exposed" or "unexposed" with respect to the risk factor under investigation. The participants are followed over time and the number of incident cases of disease determined. The incidence is calculated for the exposed and unexposed groups, and the incidence ratio (CIR or IRR) is estimated. Cohort studies can be closed, where people are enrolled only at the beginning of follow up (after which the study is closed), or open, where enrolment occurs over time.

Bowman and colleagues conducted a closed cohort study in the Gambia to examine the association between trichiasis (exposure) and corneal visual loss (outcome). At baseline 639 people with trachomatous lid scarring but without corneal visual loss were identified. Some subjects had trichiasis (exposed group) while others did not (unexposed group). After 12 years, 326 of the initial cohort were retraced. Of the 26 people with trichiasis at baseline (exposed group) two had developed corneal visual loss (CI: 2/26 or 7.7% over 12 years), compared to six of the 295 people with trachomatous lid scarring without trichiasis (unexposed group) (CI: 6/295 or 2.0% over 12 years). The CIR comparing the exposed and unexposed groups was 3.78 (95% CI: 0.80 to 17.81), showing that the 12 year risk of developing corneal visual loss was almost four times higher in people with trichiasis at baseline compared to those without (although the confidence interval included the null value).

This example illustrates the problem of loss to follow up, which is often encountered by cohort studies. Only 326 of the initial cohort of 639 were traced, and it was not known what happened to the remaining 313 people in terms corneal visual loss. Those lost to follow up often differ from those traced in terms of their exposure and disease status. Loss to follow up can therefore lead to selection bias. Another problem is competing risks, where cohort members die or develop other diseases so that they can no longer develop the disease of interest and the "true" cumulative incidence is underestimated. The problems of competing risks and loss to follow up can be overcome by using person time analyses and calculating incidence rates, as people only contribute person time while they are eligible of becoming a case. Person time analyses also allow people to switch between exposure status, contributing person time first to one exposure group and then to another. However, calculating incidence rates requires ongoing measurement of exposure and disease status, and this is logistically difficult.

The strengths of the cohort study design include the ability to measure the incidence of disease, the assurance (more or less) that the exposure preceded the disease, and the fact that several disease outcomes can be measured. The main drawback is that since most diseases are rare we either need a large sample or long follow up to accumulate enough cases to have sufficient power to make meaningful inferences. This makes cohort studies expensive and time consuming, which is why there are only a few in the ophthalmic literature.

Case-control studies

Case-control studies are used to study the aetiology of disease. Case-control studies are conducted by recruiting people who have the disease of interest (cases) as well as people without the disease (controls). The controls should be selected from the same population that gave rise to the cases (that is, if any of the controls had developed the disease they would have been included as a case) so that they represent the exposure distribution in the source population. In practice, controls are usually selected from the general population (population based case-control study) or from the same hospital as the cases (hospital based case-control study). Cases and controls are interviewed, or past medical records/laboratory files examined, to assess their exposure status. The odds of exposure in cases (number exposed versus unexposed) is compared to the odds of exposure in controls, to assess the exposure-disease association. Case-control studies are used to study the aetiology of disease. Case-control studies are conducted by recruiting people who have the disease of interest (cases) as well as people without the disease (controls). The controls should be selected from the same population that gave rise to the cases (that is, if any of the controls had developed the disease they would have been included as a case) so that they represent the exposure distribution in the source population. In practice, controls are usually selected from the general population (population based case-control study) or from the same hospital as the cases (hospital based case-control study). Cases and controls are interviewed, or past medical records/laboratory files examined, to assess their exposure status. The odds of exposure in cases (number exposed versus unexposed) is compared to the odds of exposure in controls, to assess the exposure-disease association. Case-control studies are conducted by recruiting people who have the disease of interest (cases) as well as people without the disease (controls). The controls should be selected from the same population that gave rise to the cases (that is, if any of the controls had developed the disease they would have been included as a case) so that they represent the exposure distribution in the source population. In practice, controls are usually selected from the general population (population based case-control study) or from the same hospital as the cases (hospital based case-control study). Cases and controls are interviewed, or past medical records/laboratory files examined, to assess their exposure status. The odds of exposure in cases (number exposed versus unexposed) is compared to the odds of exposure in controls, to assess the exposure-disease association. Case-control studies are conducted by recruiting people who have the disease of interest (cases) as well as people without the disease (controls). The controls should be selected from the same population that gave rise to the cases (that is, if any of the controls had developed the disease they would have been included as a case) so that they represent the exposure distribution in the source population. In practice, controls are usually selected from the general population (population based case-control study) or from the same hospital as the cases (hospital based case-control study). Cases and controls are interviewed, or past medical records/laboratory files examined, to assess their exposure status. The odds of exposure in cases (number exposed versus unexposed) is compared to the odds of exposure in controls, to assess the exposure-disease association.

Minassian and colleagues conducted a hospital based case-control study to investigate the association between childbearing and risk of cataract in young women. Cases were women aged 35–45 with bilateral "senile" cataract attending an eye hospital in central India. Controls were women of the same age with clear lenses attending the hospital with other complaints. Cases and controls were interviewed about their history of pregnancy and childbirth. The number of live births was statistically significantly higher in cases than in controls and a dose-response relation between childbearing and risk of cataract was apparent (table 2).

Case-control studies are relatively quick and cheap to carry out. They can also be used to investigate rare diseases and

| Table 2 | Odds of many versus few pregnancies in cataract cases compared to controls, hospital based case-control study | 4 |
|---|---|---|---|
| Children born | Cataract cases | Controls | Odds ratio (95% CI) |
| 1–3† | 31 | 134 | Baseline |
| 4–6† | 49 | 114 | 1.9 (1.1 to 3.1) |
| 7–11† | 15 | 14 | 4.6 (2.0 to 10.6) |

*Unexposed group, †exposed group.
multiple exposures. Recall bias is a problem, however, since cases may report exposures differently from controls. This problem is less serious when the exposure can be objectively assessed (for example, height), accurately recalled (for example, age), or verified (for example, treatment received). There is also the potential for selection bias, particularly in the selection of the controls. Case-control studies can be unfairly maligned because of these limitations, but if designed and executed thoughtfully they can provide the same information as cohort studies in a shorter time and at lower cost.

Randomised controlled trials

A randomised controlled trial (RCT) is an intervention study that forms a special subset of cohort studies.\textsuperscript{14–15} RCTs are typically used to assess the benefit of a new drug or treatment, but they are also useful for evaluating the impact of a preventive measure (for example, health education). In an RCT, people are selected and randomised to receive either the intervention (treatment under investigation) or the control (placebo or standard treatment). The purpose of randomisation is to make the intervention and control groups as similar as possible with respect to important confounders (both known and unknown) so that the only difference between them is that one group receives the intervention while the other does not. The two groups are monitored over time for the defined outcomes, allowing the relative risk to be calculated.

People will not always comply with the assigned intervention; they may be lost to follow up, refuse treatment, or else privately use different treatments. However, data analysis should only take account of the group to which they were assigned (intend to treat analyses) to avoid breaking the randomisation, as otherwise the difference between the groups could be the result of baseline differences in risk factors rather than the result of the intervention. Participants in an RCT often will not know whether they are receiving the intervention or the control (the study is blinded or masked) and ideally the investigator will also be unaware of their treatment status (the study is double blinded/masked). Blinding helps to reduce information bias.\textsuperscript{16–17}

Gordon and colleagues conducted an RCT in Cameroon to assess the effectiveness of new treatment regimens for onchocerciasis.\textsuperscript{18} They randomised 657 people with onchocerciasis to receive 150 µg/kg ivermectin yearly (standard practice group), 150 µg/kg every 3 months, 400 µg/kg then 800 µg/kg yearly, or 400 µg/kg then 800 µg/kg every 3 months. The randomisation resulted in a similar profile of baseline characteristics in the four groups, so that the only difference was the treatment assigned (table 3). Both the participants and investigators were blinded to the treatment group status. The primary outcome was the vital status of female worms which was assessed after 3 years of follow up. Significantly more female worms had died in the three monthly treated groups than in the standard practice group (OR = 2.17, 95% CI: 1.42 to 3.31 for 800 µg/kg) and 3 monthly treatment was also superior to standard practice in terms of reducing the number of female worms, itching, skin lesions, and transmission of onchocerciasis. These results are reliable because the effects of confounding, selection bias, and information bias have been minimised.

RCTs are often held up as the gold standard of study designs. They are, however, expensive, take a long time to generate results, and can only be used to answer certain types of questions. There are also ethical issues concerning the use of RCTs, as it is immoral to give a treatment that is known to be worse, or to withhold a treatment that is better, than standard practice or placebo. In designing RCTs there must therefore be a state of equipoise where the potential benefits of a new treatment are equally and oppositely outweighed by the potential harm. A committee should be established to ensure that equipoise is maintained throughout the trial.

<table>
<thead>
<tr>
<th>Table 3</th>
<th>Baseline characteristics of the participants in the four treatment arms of the onchocerciasis treatment randomised controlled trial\textsuperscript{17}</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline characteristics*</td>
<td>150 µg/kg yearly (n = 166)</td>
</tr>
<tr>
<td>Age (years)</td>
<td>39.6</td>
</tr>
<tr>
<td>Height (cm)</td>
<td>172.1</td>
</tr>
<tr>
<td>Weight (kg)</td>
<td>62.6</td>
</tr>
<tr>
<td>Palpable nodules</td>
<td>4.8</td>
</tr>
<tr>
<td>Patients with skin microfilariae (%)</td>
<td>99</td>
</tr>
<tr>
<td>Microfilariae</td>
<td>62.9</td>
</tr>
</tbody>
</table>

*Values are mean unless otherwise stated.

Box 2 Suggestions for further reading

Epidemiology methods

- Last JM, ed. A dictionary of epidemiology. New York: Oxford University Press, 2001.\textsuperscript{3} Useful as a reference for epidemiological terms, but not appropriate as an introductory text.
- Rothman KJ. Epidemiology: an introduction. New York: Oxford University Press, 2002.\textsuperscript{3} Up to date comprehensive introductory text incorporating more sophisticated concepts.

Ophthalmic epidemiology

We hope that this overview has shown you that there is no ideal study design. It is more a case of “horses for courses” as different designs are appropriate in different situations.

CONCLUSIONS
In this review we gave a whirlwind account of epidemiology as it is relates to ophthalmologists. We hope that this will increase your ability to read papers critically, and your effectiveness when communicating about epidemiology, whether orally or in writing. In one review we cannot hope to do more than touch on a number of important issues and for those interested in exploring these topics more deeply we provide a list of suggestions for further reading (box 2).

Authors’ affiliations
H Kuper, C Gilbert, International Centre for Eye Health, Clinical Research Unit, Department of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, UK

REFERENCES