What Does the Odds Ratio Estimate in a Case-Control Study?

NEIL PEARCE*†

Pearce N (Department of Medicine, Wellington School of Medicine, PO Box 7343 Wellington, New Zealand). What does the odds ratio estimate in a case-control study? International Journal of Epidemiology 1993; 22: 1189-1192.

The use of the term 'odds ratio' in reporting the findings of case-control studies is technically correct, but is often misleading. The meaning of the odds ratio estimates obtained in a case-control study differs according to whether controls are selected from person-time at risk (the study base), persons at risk (the base-population at risk at the beginning of follow-up), or survivors (the population at risk at the end of follow-up). These three methods of control selection correspond to estimating the rate ratio, risk ratio, or the odds ratio respectively, by means of calculating the odds ratio in the subjects actually studied. None of these estimation procedures depends on any rare disease assumption. Where the rare disease assumption is relevant is whether the effect which is estimated (e.g. the odds ratio) is approximately equal to some other effect measure of interest (e.g. the risk ratio or rate ratio) in the underlying study base. To avoid confusion on this issue, authors should be encouraged to not only specify the manner in which controls have been selected (e.g. by density sampling) but also the corresponding effect measure which is being estimated (e.g. the rate ratio) by the 'odds ratio' which is obtained in a case-control analysis.

In this paper I will argue that the use of the term 'odds ratio' in reporting the findings of case-control studies is technically correct, but is often misleading. I will first briefly review the commonly used measures of disease occurrence and measures of effect in cohort studies, before discussing measures of effect in case-control studies.

MEASURES OF DISEASE OCCURRENCE

Epidemiological studies should be based on the experience of a particular group of people followed over a particular period of time. Miettinen has termed this study population the 'base population' and its experience over time the 'study base'.

Table 1 shows the findings of a hypothetical cohort study of 100000 persons exposed to a particular risk factor, and 100000 people who are not exposed; both groups are followed for 10 years. For simplicity, I will assume that the outcome of interest is mortality from any cause in a fixed cohort; similar arguments apply when studying specific causes of death (or incidence of a non-fatal disease) in an open (dynamic) population, but the estimation procedures are less straightforward.

Three measures of disease occurrence are commonly used in cohort studies. These have been extensively discussed in standard texts and will only be briefly reviewed here, using the findings for the non-exposed group.

Perhaps the most common measure of disease occurrence is the (person-time) incidence rate (or incidence density) which is a measure of the disease occurrence per unit time. In this example, the non-exposed group contributed 951626 person-years during the 10 years of follow-up (this is less than the total possible person-time of 100000 person-years since people who died before the end of the 10-year period stopped contributing person-time at the time of their death) and there were 9516 deaths during the same period; thus, the incidence rate in the non-exposed group (b/Yo) was 9516/951626 = 0.0100 (or 1000 per

<table>
<thead>
<tr>
<th></th>
<th>Exposed</th>
<th>Non-exposed</th>
<th>Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deaths</td>
<td>18 127 (a)</td>
<td>9516 (b)</td>
<td></td>
</tr>
<tr>
<td>Survivors</td>
<td>81 873 (c)</td>
<td>90 484 (d)</td>
<td></td>
</tr>
<tr>
<td>Base population</td>
<td>100 000 (N,)</td>
<td>100 000 (N,)</td>
<td></td>
</tr>
<tr>
<td>Person-years</td>
<td>906 346 (Y,)</td>
<td>951 626 (Yo)</td>
<td></td>
</tr>
<tr>
<td>Incidence rate</td>
<td>0.0200 (I,)</td>
<td>0.0100 (Io)</td>
<td>2.00</td>
</tr>
<tr>
<td>Cumulative incidence</td>
<td>0.1813 (CI,)</td>
<td>0.0952 (CIO)</td>
<td>1.90</td>
</tr>
<tr>
<td>Incidence odds</td>
<td>0.2214 (O,)</td>
<td>0.1052 (Oo)</td>
<td>2.11</td>
</tr>
</tbody>
</table>

* Department of Medicine, Wellington School of Medicine, PO Box 7343, Wellington, New Zealand.
† Formerly at IARC, 150 cours Albert-Thomas, 69372 Lyon, France.
100,000 person-years). The incidence rate is a rate per
unit time, and has the reciprocal of time as its dimen-

sion. A second measure of disease occurrence is the
cumulative incidence (incidence proportion) which is
the proportion of study subjects who experience the
outcome of interest at any time during the follow-up
period. In this instance, there were 9516 deaths among
the 100,000 people in the non-exposed group, and the
cumulative incidence (b/N₀) was therefore 9516/
100,000 = 0.0952. The cumulative incidence is a pro-
portion and is dimensionless, but it is necessary to
specify the time period over which it is being
measured. When the outcome of interest is rare over
the follow-up period, then the cumulative incidence is
approximately equal to the product of the incidence
rate times the follow-up period (in this instance this
product is 0.1000 whereas the cumulative incidence is
0.0952).

A third measure of disease occurrence is the inci-
dence odds which is the ratio of the number of sub-
jects who experience the outcome (b) to the number of
subjects who do not experience the outcome (d). In
this instance, the incidence odds (b/d) is 9516/90,484
= 0.1052. As for the cumulative incidence, the in-
cidence odds is dimensionless but it is necessary to
specify the time period over which it is being
measured. When the outcome is rare over the follow-
up period then the incidence odds is approximately
equal to the cumulative incidence.

These three measures of disease occurrence all in-
volve the same numerator: the number of deaths (b).
They differ in whether their denominators represent
person-time at risk (Y₀), persons at risk (N₀), or sur-
vivors (d).

MEASURES OF EFFECT IN COHORT STUDIES
Corresponding to these three measures of disease oc-
currence, there are three principal multiplicative
measures of effect which are used in cohort studies.
The measure of primary interest is often the rate
ratio (incidence density ratio) which is the ratio of the
incidence rate in the exposed group (a/Y₁) to that in
the non-exposed group (b/Y₀). In the example in Table
1, the incidence rates are 0.02 per person-year in the
exposed group and 0.01 per person-year in the non-
exposed group, and the rate ratio is 2.00.

A second effect measure is the risk ratio (cumulative
incidence ratio) which is the ratio of the cumulative
incidence in the exposed group (a/N₁) to that in the non-
exposed group (b/N₀). In this example, the cumulative
incidence ratio is 0.1813/0.0952 = 1.90. When the
outcome is rare over the follow-up period the risk
ratio is approximately equal to the rate ratio.

A third possible effect measure is the (incidence) odds ratio which is the ratio of the incidence odds in
the exposed group (a/c) to that in the non-exposed
group (b/d). In this example the odds ratio is
0.2214/0.1052 = 2.11. Once again, when the outcome
is rare over the study period the incidence odds ratio is
approximately equal to the incidence rate ratio.

Each of these effect measures involves the ratio of a
measure of disease occurrence in the exposed group to
that in the non-exposed group. The various measures
disease occurrence all involve the same numerators
(deaths), but differ in whether their denominators are
based on person-time, persons, or survivors. They are
all approximately equal when the disease is rare during
the follow-up period (e.g. a cumulative incidence of
less than 10%). The rate ratio is sometimes regarded as
the primary effect measure on theoretical grounds, but
the risk ratio is perhaps the easier to conceptualize,
especially by non-epidemiologists, whereas the odds
eratio has been severely criticized as an effect
measure. These three multiplicative effect measures
are sometimes referred to under the generic term of
relative risk.

MEASURES OF EFFECT IN CASE-CONTROL
STUDIES
Suppose that a nested case-control study is conducted
in this study base, involving all of the deaths and a
group of controls. The effect measure which this case-
control study will estimate depends on the manner in
which controls are selected. Once again, there are three
main options.

One option is to select controls from those who do
not experience the outcome during the follow-up
period, i.e. the survivors (at the end of follow-up). In
this instance, a sample of controls chosen from the sur-
vivors will estimate the exposure odds (b/d) of the sur-
vivors, and the odds ratio obtained in the case-control
study will therefore estimate the odds ratio in the base
population. Early presentations of the case-control ap-
proach were often presented in this context, and it was
emphasized that the odds ratio was approximately
equal to the risk ratio when the disease was rare.

It was later recognized that controls can be sampled
from the entire base population (those at risk at the
beginning of follow-up), rather than just from the sur-
vivors (those at risk at the end of follow-up). This ap-
proach which was previously used by Thomas and
Kupper et al, has more recently been termed ‘case-
base’ sampling, or the ‘case-cohort’ design. In this
instance, the controls will estimate the exposure odds in the base population of persons at risk at the start of follow-up (N_e/N_o), and the odds ratio obtained in the case-control study will therefore estimate the risk ratio in the base population (in this instance the method of calculation of the odds ratio is the same as for any other case-control study, but minor changes are needed in the standard methods for calculating confidence intervals and P values to take into account that some cases may also be selected as controls)\(^5\).

The third approach is to select controls longitudinally throughout the course of the study;\(^5,13\) this is sometimes described as 'risk-set sampling',\(^14\) 'sampling from the study base' (the person-time experience),\(^3\) or 'density sampling'.\(^15\) In this instance, the controls will estimate the exposure odds in the study base (i.e. the person-time at risk, which can be conceptualized as 'person-years', 'person-months' or even 'person-days'), and the odds ratio obtained in the case-control study will therefore estimate the rate ratio in the study base. Although case-control studies have traditionally been presented in terms of sampling from the survivors,\(^7\) it has been pointed out\(^4\) that most case-control studies actually involve density sampling (with matching on a time variable such as calendar time or age), and therefore estimate the rate ratio without the need for any rare disease assumption\(^5,13,16\) (it should be noted that density sampling does not involve a random sample of the person-time in the study base since controls are only sampled for the 'instantaneous' time periods in which cases occur; thus the odds ratio obtained from density sampling may not be the same as that obtained by selecting controls at random from the study base, but the two odds ratios will be equivalent if 'time' is controlled in the analysis—in this instance 'density sampling' involves matching and then controlling for time in the analysis whereas 'random sampling' can be followed by directly controlling for time in the analysis).

DISCUSSION

Thus, the meaning of the odds ratio estimates obtained in a case-control study differs according to whether controls are selected from person-time at risk (the study base), persons at risk (the base-population at risk at the beginning of follow-up), or survivors (the population at risk at the end of follow-up). These three methods of control selection correspond to estimating the rate ratio, risk ratio, or the odds ratio respectively, by means of calculating the odds ratio in the subjects actually studied. None of these estimation procedures depend on any rare disease assumption. Where the rare disease assumption is relevant is whether the effect which is estimated (e.g. the odds ratio) is approximately equal to some other effect measure of interest (e.g. the risk ratio or rate ratio) in the underlying study base.

Thus, the use of the term 'odds ratio' in reporting the findings of case-control studies is technically correct, since it does involve calculating the odds ratio in the subjects actually studied. However, the traditional presentation of case-control studies and the universal use of the term 'odds ratio', has led to some confusion as to what is being estimated, whether a rare disease assumption is required, and whether the 'odds ratio' is a measure of intrinsic interest.

One solution would be to define a new term to refer specifically to the odds ratio in a case-control study (e.g. 'case-control odds ratio'). This would perhaps be inadvisable, given the exponential increase in new names for old concepts in epidemiology in recent years. Nevertheless, the situation would be clarified if authors were encouraged to not only specify the manner in which controls have been selected (e.g. by density sampling) but also the corresponding effect measure which is presumably being estimated (e.g. the rate ratio) by the 'odds ratio' which is obtained in a case-control analysis.

ACKNOWLEDGEMENTS

This work was conducted in part during the tenure of a Senior Research Fellowship of the Health Research Council of New Zealand, and in part during the tenure of a Visiting Scientist Award of the International Agency for Research on Cancer. I am grateful to Sander Greenland and David Savitz for their comments on the draft manuscript.

REFERENCES


(Revised version received June 1993)