A recent debate in this\textsuperscript{1–3} and other journals\textsuperscript{4–9} has pointed out the interest surrounding cross-sectional studies and the epidemiological measures used to convey their results. Some authors (see for example\textsuperscript{1,2,7,8}) have indicated their preference for the use of prevalence rate ratios (PRR) against the more frequently encountered prevalence odds ratios (POR). Others\textsuperscript{3} have claimed the utility of both depending on many arguments and/or circumstances.

It is likely that the great availability of computer programs for the estimation of PRR has probably reduced their application. This gap is likely to reduce\textsuperscript{13,14} and the selection of appropriate epidemiological measures should not be based on the available tools but on epidemiological grounds.

To add further substance to the discussion, and expand on some suggestions made by other authors in this Journal\textsuperscript{3} we thought it useful to clarify the mathematical relationship between PRR and POR and to evaluate the degree of divergence of the two measures as a function of the prevalence of disease and exposure.

METHODS

If we consider a $2 \times 2$ table deriving from a cross-sectional study a number of useful quantities can be easily defined, including: prevalence of the disease ($Pr(D)$), prevalence of the exposure ($Pr(E)$), prevalence odds ratio ($POR$), and prevalence rate ratio ($PRR$) (Table 1). With the help of some algebra different
relationships between PRR and POR can be established, and a useful resulting formula is the following:

\[ \text{POR} = \frac{\text{PRR}}{1 - \text{Pr}(E) + \text{PRR} \times \text{Pr}(E) - \text{PRR} \times \text{Pr}(D)} \]

The selected formulation enables us to explore the relationship between POR and PRR as a function of both the prevalence of the disease and the prevalence of the exposure. Implicitly, this choice corresponds to evaluating how the POR departs from the PRR which is conceptually taken as a reference measure: if PRR as a function of POR is of interest, a reverse approach can be easily developed.

\( \text{Pr}(D) \) and \( \text{Pr}(E) \) could range from 0 to 1 while POR and PRR could extend from 0 to infinity, but due to the relationships between these measures some combinations of the values are not permitted because they give rise to, for example, negative or undefined quantities.

RESULTS

A simple look at formula (1), with a little algebra, shows that when PRR is equal to one POR will coincide exactly with PRR irrespective of the values of \( \text{Pr}(D) \) and \( \text{Pr}(E) \), while for all other conditions POR and PRR will differ. In addition, when PRR is greater than one POR will be greater than PRR, while when PRR is less than one POR will be less than PRR: both these departures from equality greatly depend on the values of \( \text{Pr}(D) \) and \( \text{Pr}(E) \).

Figure 1 shows the relationship between PRR and POR for selected values of the prevalence of the disease and of the exposure, in a restricted range of PRR values. The figure highlights that the relationship is not linear (it is a quadratic curve) and that according to the values of \( \text{Pr}(D) \) and \( \text{Pr}(E) \) the curve rotates around the value of

![Graphical relationship between prevalence rate ratio (PRR) and prevalence odds ratio (POR) for selected values of the prevalence of the disease (Pr(D)) and the prevalence of the exposure (Pr(E)).](image)
PRR equal to one and/or changes curvature. In addition, for the special case of $Pr(D) = Pr(E) = 0.5$ the POR value is exactly the square of the corresponding PRR.

If we consider for the sake of comparison as a baseline the curve which corresponds to a value of $Pr(D) = Pr(E) = 0.5$ (black triangles) Figure 1 shows that a decrease in the prevalence of the exposure (e.g. $Pr(E) = 0.2$, black rectangles) will cause an increase in the curvature of the relationship giving rise to values of POR which are always greater than the baseline; in particular they will depart much more from the PRR values when PRR is greater than one and will approach the PRR values when PRR is less than one.

If an increase in the prevalence of the exposure is considered (e.g. $Pr(E) = 0.8$, white circles) a decrease in the curvature of the relationship will result, with POR departing more from the PRR values when PRR is $<1$ and approaching them when PRR is $>1$.

A different result will be obtained if a change in the prevalence of the disease (instead of a change in the prevalence of the exposure) is considered: in this situation a rotation of the curve around the point $POR = PRR = 1$ will take place. For example, again with respect to the curve with $Pr(D) = Pr(E) = 0.5$, if we consider a decrease of $Pr(D)$ (e.g. $Pr(D) = 0.2$, white rectangles) the POR will always be closer to the corresponding PRR value, while an increase in the prevalence of the disease (e.g. $Pr(D) = 0.8$, black circles) will cause a further departure of the POR from PRR: both the changes do not depend on PRR being greater or less than one.

In addition, a change in the prevalence of the disease will affect the POR value much more than a change in the prevalence of the exposure, as can be seen from comparison of the white and black rectangle curves in Figure 1 (or the white and black circle curves).

When the disease is rare ($Pr(D) < 0.10$) no major discrepancies emerge between POR and PRR, irrespective of the prevalence of the exposure. For example, with $Pr(D) = 0.05$ and $PRR = 2.5$ the values of POR range from 2.71 when $Pr(E) = 0.01$ to 2.58 when $Pr(E) = 0.99$, which are not very different from 2.5, and the differences tend to diminish as PRR approaches one.

**DISCUSSION**

Many epidemiological measures can usefully describe the results of a study, and the selection of the most appropriate is never obvious. In particular, some authors have pointed out that the odds ratio is a particularly misunderstood measure and that the cross-sectional study requires extensive methodological discussion.

Cross-sectional data can serve many purposes, and the wide range of applications could suggest the use of different epidemiological measures in different contexts.

For example, cross-sectional data can be used to estimate incidence density ratios (IDR). In this situation it has been shown that under some restrictive assumptions (e.g. the distribution over time of exposures, covariates, and incidence; migration among diseased; duration of disease) POR approximate IDR better than PRR, which means that when we are dealing with chronic diseases (i.e. long latency diseases with different follow-up periods for the subjects under observation) the use of POR may be completely justified.

On the other hand, if risk ratios (RR) are the parameters of interest (considering acute diseases, with follow-up periods similar among subjects) then PRR should be the measure of choice.

In other applications we are interested in outcomes which are not strictly ‘diseases’: consider, for example, a study with the aim of describing how the proportion of subjects with a particular condition (e.g. an adducts level above the median) varies according to some covariate(s). In these situations the use of the prevalence rate (and hence of PRR) as a descriptor of a ‘state’ seems more natural and intelligible measure.

In addition, it has been noted that the usual assumption of similar duration of disease (or prognosis) between exposed and unexposed subjects may not be satisfied: a case in point are musculoskeletal disorders for which duration of symptoms are likely to vary between exposure groups. In this situation again the use of PRR seems warranted.

A further argument against the use of POR is that it can introduce confounding even when there is none in terms of prevalence rates.

In other instances (the ‘sex ratio’, for example) the odds ratio is a natural epidemiological measure. As a further step in the discussion we have explored the relationship between POR and PRR in order to understand the most important discrepancies between them, using a general formula for this relationship as a function of the prevalence of the disease ($Pr(D)$) and the prevalence of the exposure ($Pr(E)$). The results indicate that the POR is always further away from the null value than the PRR and that the discrepancies between POR and PRR strongly depend on both $Pr(D)$ and $Pr(E)$, with the former being more important from a quantitative point of view.

With respect to the prevalence of the disease we have considered values around 0.5 because they are very common in emerging areas like musculoskeletal disorders and molecular epidemiology. In the latter, for
example, the outcome may be represented by a categorization of a continuous variable, e.g. adducts level or other biological markers, with the median value being the cutpoint. In other areas the most frequent prevalence value of the disease and of the exposure can greatly vary from the presented examples and consequently the discrepancies between POR and PRR could vary as well. In particular, it is well known that when $Pr(D)$ is low (less than, say, 0.10) POR and PRR will be very similar and there will be no practical reasons to distinguish between them.

We have chosen to describe the relationship between POR and PRR in terms of the prevalence of the disease instead of, for example, the prevalence of the disease among non-exposed subjects because in many cross-sectional studies the exposure status is not a criterion for selecting subjects into the study. We have only addressed point estimates of POR and of PRR. Statistical aspects, like test of hypothesis or confidence interval estimation, can easily be included in the discussion recalling that both measures are defined in the frame of binomial variability. From the point of view of hypothesis testing (i.e. accepting/rejecting the hypothesis that POR or PRR is equal to one) the two measures give in practice the same answer, whereas in terms of confidence intervals POR is characterized by wider intervals which, in other words, means less precision in the estimates.

The issue of the discrepancy in point estimation requires a further comment. When POR and PRR are considered in their own domain, the two measures do not need to be compared, but when they are interpreted as estimators of some ‘risk’ (ratio of risks) a new problem arises. Suppose, for example, that PRR is equal to two and POR is equal to four (which is the case when $Pr(D) = Pr(E) = 0.5$). Irrespective of their statistical significance (confidence intervals), the two values indicate, from a quantitative point of view, very different perspectives of interpretation (a twofold versus a fourfold risk) which depend only on the choice of a specific epidemiological measure and not on the scientific question at issue. We should be aware of this bias, particularly when comparing results of different studies from the literature.

In summary, irrespective of the preference of different authors or the habits induced by software tools availability, we think that POR and PRR have their specific role with cross-sectional data and that the choice between them should remain on epidemiological grounds only.

The arguments made above would suggest the use of PRR instead of POR in most situations. Despite these considerations POR have played a major role in the description of the results of cross-sectional studies due mainly to mathematical convenience and the easy availability of advanced statistical tools (logistic regression, mainly) which were not so easy to use for the estimation of PRR. This situation is expected to change rapidly thanks to new software programs (e.g. SAS GENMOD).

REFERENCES


(Revised version received July 1996)