Biostatistics Series Module 8: Assessing Risk

Avijit Hazra, Nithya Gogtay

Abstract

In observational studies, as well as in interventional ones, it is frequently necessary to estimate risk that is the association between an observed outcome or event and exposure to one or more factors that may be contributing to the event. Understanding incidence and prevalence are the starting point in any discussion of risk assessment. Incidence rate uses person-time as the denominator rather than a simple count. Ideally, rates and ratios estimated from samples should have their corresponding 95% confidence intervals (CIs). To assess the importance of an individual risk factor, it is necessary to compare the risk of the outcome in the exposed group with that in the nonexposed group. A comparison between risks in different groups can be made by examining either their ratio or the difference between them. The $2 \times 2$ contingency table comes in handy in the calculation of ratios. Odds ratio (OR) is the ratio of the odds of an event in the exposed group, to the odds of the same event in the nonexposed group. It can range from zero to infinity. When the odds of an outcome in the two groups are identical, then the OR equals one. OR >1 indicates exposure increases risk while OR <1 indicates that exposure is protecting against risk. The OR should be presented with its 95% CI to enable more meaningful interpretation – if this interval includes 1, then even a relatively large OR will not carry much weight. The relative risk (RR) denotes the ratio of risk (probability) of event in exposed group to risk of same event in the nonexposed group. Its interpretation is similar (but not identical) to the OR. If the event in question is relatively uncommon, values of OR and RR tend to be similar. Absolute risk reduction (ARR) is a measure of the effectiveness of an intervention with respect to a dichotomous event. It is calculated as proportion experiencing the event in control group minus the proportion experiencing the event in treated group. It is often used to denote the benefit to the individual. The reciprocal of ARR is the number needed to treat (NNT), and it denotes the number of subjects who would need to be treated to obtain one more success than that obtained with a control treatment. Alternatively, this could also denote the number that would need to be treated to prevent one additional adverse outcome as compared to control treatment. Extended to toxicity, the NNT becomes a measure of harm and is then known as the number needed to harm (NNH). NNT and NNH are important concepts from the policy makers perspective and ideally should be calculated in all trials of therapeutic or prophylactic intervention.

Key Words: Absolute risk reduction, attributable fraction, attributable risk, incidence, incidence rate, number needed to harm, number needed to treat, odds ratio, prevalence, rate ratio, relative risk, relative risk reduction

Introduction

Risk estimation is part of our daily lives. Risks associated with particular actions, events, or exposures are communicated all the while through mass media and scholarly publications. In the health-care setting, risk measures are communicated to physicians through research publications and reviews, who in turn will have to integrate them in the clinical decision-making process. The risk estimates will also need to be communicated in comprehensible terms to patients, caregivers, and study volunteers.

Various risk estimates have evolved to quantify the risk associated with particular exposures toward the occurrence of disease or adverse events. To appropriately interpret and apply these estimates, it is first necessary to understand them. However, the multiplicity of estimates appears confusing to the uninitiated. Furthermore, the

How to cite this article: Hazra A, Gogtay N. Biostatistics series module 8: Assessing risk. Indian J Dermatol 2017;62:123-9.
Received: February, 2016. Accepted: February, 2016.
use of different estimates at times seems to convey different levels of risk or even contradictory ideas of whether risk exists at all. In the face of such conflicting information pertaining to risk, it may be quite difficult for both physicians and patients to use risk estimates for decision-making. Therefore, a clear understanding of the statistics of risk estimates is important.

Incidence and Prevalence

Although these are not risk estimates per se, they are the starting points in any discussion on risk assessment.

\[
\text{Incidence} = \frac{\text{Number of new cases of a disease or event occurring in a specified time period}}{\text{Number of subjects at risk of the disease or event during the same period}}
\]

\[
\text{Prevalence} = \frac{\text{Total number of affected subjects in a population in a specified time period}}{\text{Number of subjects in the population at the time}}
\]

Incidence and prevalence are both rates commonly used in epidemiological studies and are easier to understand and compare when projected to common denominators such as per 1000 or per 100,000 population.

Suppose in a community of 25,000 individuals, there are already 16 diagnosed cases of cutaneous malignant melanoma, and 4 new cases have been detected in 2016. Thus, at the end of 2016, there are twenty diagnosed cases. The prevalence is therefore 20/25000 that is 0.8/1000 population. The number of individuals at risk in 2016 is 24,984 since 16 already have the disease. Thus, the incidence is 4/24,984 that is 0.16/1000 population. Note that if the absolute numbers are low compared to the size of the population, we can take the population size (rather than the numbers at risk) as the denominator for calculation of incidence. Thus, the incidence may be taken as 4/25,000, which is still 0.16/1000.

The prevalence is most useful for measuring the burden of chronic disease in populations rather than acute incidents or illnesses. Epidemiologists use some variations of the basic prevalence (number of all cases divided by population size) concept. Point prevalence is the proportion of a population that has the condition at a specific point in time. Period prevalence is the proportion of a population that has the condition at some time during a given period (e.g., over a year) and includes people who already have the condition at the start of the period as well as those who acquire it during that period. Lifetime prevalence is the proportion of a population that at some point in their life (up to the time of assessment) has experienced the condition.

If the incidence of a disease fluctuates over time, we can calculate something-like an average annual incidence. This is simply the incidence calculated individually over a certain number of years and then averaged out over the period. Some studies require the calculation of an incidence rate or incidence density which is:

\[
\text{Incidence rate} = \frac{\text{Number of new cases}}{\text{Sum of person – Time at risk}}
\]

Incidence rate or density can be measured in a closed cohort or in an open population. Its numerator is the same as in calculating basic incidence; however, the denominator is calculated differently.

In case of an open population, the sum of person-time at risk is approximately equal to \( N \times D \), where \( N \) is average population size and \( D \) is the duration of study. For example, a population with an average size of 1000 studied for 1 year accounts for 1000 person-years while the same population studied for 3 years accounts for 3000 person-years.

In case of a closed cohort, we can either count person-time individually for each subject in the cohort and sum or break the cohort into those who remain healthy (Group 1) and those who develop disease (Group 2) and calculate (an approximate person-time at risk) as follows:

- For Group 1, let \( N_1 \) be the size and \( D_1 \) be the duration of observation; the person-time \( T_1 \) is then \( N_1 \times D_1 \)
- For Group 2, let \( N_2 \) be the size and \( D_2 \) be the duration of observation; the person-time \( T_2 \) is then \( N_2 \times 0.5 \times D_2 \)
- The total person-time at risk is then \( T_1 + T_2 \).

Relation between Incidence and Prevalence

Although a higher incidence is expected to contribute to a higher prevalence, the actual relationship between incidence and prevalence depends greatly on the natural history of the disease or event in question. For instance, during an influenza epidemic, the incidence will be high but will not contribute to much growth of the prevalence because of the high rate of spontaneous resolution of disease. If mortality in an infectious disease is high, such as in Ebola, then also the prevalence will be relatively low. In case of a disease that may not be curable, but where treatment permits sustained survival, such as diabetes, the incidence will contribute to continuous growth of prevalence. However, the mortality which occurs in the population will limit the prevalence. Obviously, the prevalence will continue to grow until mortality equals or exceeds the incidence.

If the population is initially in a “steady state” that is prevalence is fairly constant and incidence and...
outcome (cure or death) are about equal, then the relationship among these three parameters is:

\[
PR/(1 - PR) = IR \times D
\]

Where \( PR \) = proportion of the population with disease, \((1 - PR)\) is proportion without disease, \( IR \) is the incidence, and \( D \) is average duration of disease that is the average time from diagnosis till cure or death. If the disease is uncommon (say affecting <10% of the population), then the relationship simplifies to:

\[
PR = IR \times D
\]

If the average duration of disease remains constant, then preventive measures that reduce the incidence of disease should lead to decreasing prevalence. If incidence remains constant, then development of a curative treatment would reduce the average duration of disease and thereby also reduce prevalence. On the other hand, introduction of a noncurative treatment combined with effective preventive measures would lead to an apparently contradictory situation where incidence falls, but prevalence keeps rising since the average duration of the disease is lengthened. This is happening with the AIDS epidemic in India after the introduction of effective control measures and wide availability of antiretroviral therapy.

This relationship can also be used to estimate the average duration of disease in the “steady state” situation. For instance, if the incidence of gastric cancer is 52 new cases/100,000 population and the prevalence is 26/100,000, then average disease duration would be 0.5 year. Thus, gastric cancer patients are expected to live 6 months on an average from the time of diagnosis till death.

Note that since incidence and prevalence are usually based on studies with samples, whereas we really need estimates pertaining to underlying populations, they should ideally be expressed with their 95% confidence intervals (CIs) to provide an idea of the population figures.

The 95% CI for a proportion is given by:

\[
p \pm 1.96 \times \text{Standard error of proportion}
\]

\[
p \pm 1.96 \times \sqrt{(p [1 - p]/n)}
\]

Where \( p \) is the sample proportion, and \( n \) is the sample size.

Note that the calculation of standard error of proportion as square root of \((p [1 - p]/n)\) is actually an approximation for situations where the expected population size is considerably larger (at least twenty times) than the sample size. If this condition is not met, then there are other versions of the formula to be used. In addition, the calculation assumes that the sampling has been random and large enough to represent the population adequately.

Risk Estimates as Measures of Treatment Difference

For continuous data, the difference between two treatments is usually summarized as the difference between two means. For example:

- Difference between atorvastatin and rosuvastatin on mean low-density lipoprotein (LDL)-cholesterol level after 24 weeks
- Difference between atorvastatin and rosuvastatin on mean reduction in LDL-cholesterol level after 24 weeks of therapy.

For categorical data where the outcome is absolute (response/nonresponse), there are a number of ways to summarize a treatment difference such as absolute risk reduction (ARR), relative risk reduction (RRR), and number needed to treat (NNT). If we look at the formulas below, their calculation is straightforward. However, they need to be interpreted carefully, and the choice of the summary measure used can affect how the treatment difference is interpreted.

\[
ARR = (% \text{ Outcome in arm A}) - (% \text{ Outcome in arm B})
\]

\[
RRR = 1 - \frac{\text{Relativerisk}}{1 - \frac{(% \text{ Outcome in arm A})}{(% \text{ Outcome in arm B})}}
\]

\[
NNT = \frac{1}{ARR}
\]

Let us try to understand these estimates with an example. Suppose that in a parallel group randomized controlled trial (RCT) of two antibiotics in adults with cellulitis, 20% of the participants in the control treatment group experienced a bad outcome, compared to 12% in the new treatment group. Without knowing the full effects of therapy, it still appears that the new antibiotic reduces more of the bad outcome of the disease. But how meaningful is the effect?

ARR is the most straightforward way of presenting study results to enable decision-making. In this case, ARR is 0.2–0.12 i.e., 0.08% or 8%. Thus, compared to the control treatment, there is a reduction of the absolute risk by 8%. This means that, if 100 subjects were treated, 8 would be prevented from developing bad outcomes. Another way of expressing this is the NNT. If 8 subjects out of 100 benefit from treatment, the NNT for one subject to benefit is 1/0.08, i.e., 12.5. Thus, we can state that about 13 subjects would need to be on the new treatment rather than on the control treatment to avoid one instance of bad outcome. Although this is not an intuitive estimate, a little thought should convince one that this is a really useful measure for policy planners, particularly when prioritizing allocation of scarce budgetary resources. It is helpful to keep in mind that while both ARR and NNT values vary according to the absolute risk level, the former measures the benefit
to the individual while the latter measures the benefit of treating the population.

If we are simply interested in estimating the benefit of treatment without regard to the absolute risk level, we can calculate RRR. In the above example, the risk of bad outcome is reduced by the new antibiotic. The risk of bad outcome in the new treatment group compared to the control group is 0.12/0.2, i.e., 0.6. Therefore, the RRR is 1–0.6, i.e., 0.4%, or 40%. In other words, use of the new antibiotic reduces the risk of bad outcome by 40% compared to the control treatment. Note that it is a little tricky to depend only on the RRR alone to estimate benefit of treatment. Consider Table 1.

Thus, the relative risk (RR), and therefore the RRR, may be the same (33% in this case) irrespective of the absolute level of risk, whereas the ARR will be greatest in those at greatest risk. The greater the risk, the more is the benefit from the intervention. This is not reflected in RRR. Unfortunately, the RRR estimate is often used to convey the impression of a large benefit from treatment when the absolute benefit may actually be small. One should be particularly wary of RRR figures presented in the pharmaceutical promotional literature.

**Attributable Risk**

Attributable risk (ATR) is the extent of disease incidence that can be linked to a specific exposure. It is calculated in a manner analogous to ARR but is used in the context of epidemiological studies:

\[
\text{ATR} = \text{Risk (i.e., Proportion with an outcome) in exposed group} - \text{Risk (i.e., Proportion with the outcome) in nonexposed group}
\]

The difference is thus the extent of the incidence that may be attributed to the particular exposure; the incidence in the nonexposed being the background risk.

ATR is often expressed as a percentage of the risk in the exposed, and this is also referred to as the proportional ATR or the attributable fraction (AF):

\[
\text{AF} = \frac{\left(\text{Risk in exposed}\right) - \left(\text{Risk in non-exposed}\right)}{\text{Risk in exposed}} \times 100
\]

In general, all individuals, whether or not they have been exposed to a risk factor, have some chance of developing a disease if no prevention measures have been taken. AF is an useful indicator of the potential for prevention. However, this applies to only those exposed to the particular risk factor in question. If the exposure is uncommon, then, even if AF is high, the potential for prevention in the general population would be limited. The population ATR (PAR), also called the population AF provides an index of the potential for prevention at the population level, taking into consideration the prevalence of the exposure as well as the associated RR. It is calculated as:

\[
\text{PAR} = \frac{\left(\text{Risk in total population}\right) - \left(\text{Risk in exposed population}\right)}{\left(\text{Risk in total population}\right)}
\]

Which is equivalent to:

\[
\text{PAR} = p \left(\frac{RR - 1}{RR}\right)
\]

Where P is the prevalence of the exposure in the general population and RR is the relative risk associated with the exposure.

This may also be expressed as a percentage.

PAR thus provides an estimate of the burden of a disease in a population associated with a particular exposure. Further, if the impact of an intervention in reducing exposure is known (r%), then the population impact can be calculated with its help as:

\[
\text{Population impact of intervention} (%) = \text{PAR} \times r \%
\]

**More about Number Needed to Treat**

The NNT is a population level measuring the effectiveness of treatment when the outcome is dichotomous. It denotes the number of subjects who would need to be treated to obtain one more success than that obtained with a control treatment. Alternatively, this could also denote the number that would need to be treated to prevent one additional adverse outcome as compared to the control treatment.

Thus, NNT of five for a new antibiotic would indicate that five subjects would need to be treated to achieve one additional cure of infection in comparison to the control antibiotic. Alternatively, it could indicate that five subjects would need to be treated to avoid one additional death from infection in comparison to the control antibiotic. The second statement highlights the relationship of NNT with ARR, NNT being simply the reciprocal of ARR.

As we have already said NNT, although not an intuitive measure of risk communication to the individual patient, is an useful measure of treatment effectiveness from the point of view of the policy maker. For a therapeutic

---

**Table 1: Relation between risk estimates**

| Percent of bad outcome with control intervention | Percent of bad outcome with test intervention | RR | RRR | ARR (%) | NNT |
|-----------------------------------------------|---------------------------------------------|-----|-----|---------|-----|
| 60                                            | 40                                          | 0.67| 0.33| 20      | 5   |
| 40                                            | 27                                          | 0.67| 0.33| 13      | 8   |
| 30                                            | 20                                          | 0.67| 0.33| 10      | 10  |
| 15                                            | 10                                          | 0.67| 0.33| 5       | 20  |

RRR: Relative risk reduction, RR: Relative risk, ARR: Absolute risk reduction, NNT: Number needed to treat
intervention to be acceptable, its NNT needs to be small; the smaller the NNT, the more successful the intervention. Conversely, if the intervention has no effect, then NNT would be infinitely large because there would be zero risk reduction in its use.

However, in a prophylactic situation, the difference between the control and the prophylactic intervention could be small, which would result in the NNT being high, but the intervention could still be considered successful and cost-effective if the condition is serious. This is commonly the case in the evaluation of a vaccine or a drug to prevent death or serious complication.

In an RCT, the NNT of the active treatment can be readily calculated from the formula:

\[
NNT = \frac{1}{ARR} = \frac{1}{(\frac{TAR_{act}}{TOT_{act}}) - (\frac{TAR_{con}}{TOT_{con}})}
\]

Where,

\[
TAR_{act} = \text{Number given active treatment achieving target endpoint,}
\]
\[
TOT_{act} = \text{Total number given active treatment,}
\]
\[
TAR_{con} = \text{Number given control treatment achieving target endpoint,}
\]
\[
TOT_{con} = \text{Total number given control treatment.}
\]

A negative NNT value could indicate that the intervention is not achieving the desired outcome in comparison to the control treatment; in fact, it may be causing harm. Alternatively, the NNT concept may be applied to an adverse effect of a treatment. It is then referred to as the number needed to harm (NNH).

The Relative Risk and the Odds Ratio

When comparing two independent proportions by a test such as the Chi-square test or the Fisher’s exact test, \( P < 0.05 \) only indicates that there is a statistically significant difference between the two proportions. If the two proportions represent the incidence of a dichotomous event in exposed and nonexposed groups, the \( P \) value will not indicate how much risk is conferred by the exposure that information comes through calculation of the odds ratio (OR) or the RR.

The 2 × 2 contingency table comes in handy in the exploration of these ratios and we will be using it as follows:

| Group     | Outcome | Row total |
|-----------|---------|-----------|
|           | Present | Absent    |            |
| Exposed   | A       | b         | a + b      |
| Nonexposed| C       | d         | c + d      |
| Column total | a + c     | b + d     | N=a + b + c + d |

The RR, which in this context has also been called the risk ratio, estimates the increase in risk of an (adverse) outcome or event when exposed to a risk factor compared to when not exposed to the same factor. It is calculated as:

\[
RR = \frac{\text{Risk in exposed group}}{\text{Risk in non-exposed group}}
\]

\[
= \frac{a (a + b)}{c (c + d)}
\]

Note that when the risk of an outcome in the two groups is identical, the RR equals one. RR >1 indicates exposure increases risk while RR <1 indicates that exposure is reducing risk and is therefore actually a protective exposure.

Odds refer to the ratio of the probability of occurrence of an event to the probability of nonoccurrence of an event. If we roll a fair dice, the probability of getting a six is 1/6, and the probability of not getting a six is 5/6. Thus, the odds of getting a six is (1/6)/(5/6); i.e., 1/5. OR is the ratio of the odds for a dichotomous outcome between two groups. In the context of risk assessment, it is calculated as:

\[
OR = \frac{\text{Odds in exposed group}}{\text{Odds in non-exposed group}}
\]

\[
= \frac{a (a + b) / (b (a + b))}{c (c + d) / (d (c + d))} = \frac{a / b}{c / d} = \frac{ad}{bc}
\]

OR is thus the ratio of a ratio. From the derivation, it is evident that the OR can be simply calculated as the cross-products ratio from a two-dimensional contingency table.

The value of the OR can range from zero to infinity. When the odds of an outcome in the two groups are identical, the OR will be one. OR >1 indicates exposure increases risk while OR <1 indicates that exposure is reducing risk.

The RR and OR are universally used measures in epidemiological studies. They are calculated in all types of analytical observational studies (those with a control group) except in longitudinal studies with incomplete follow-up. However, risk cannot be estimated directly in a case–control study, because the size of the denominator from which cases are being drawn is unknown, and therefore, RR is not calculated for this type of study. Instead, the OR can be used to give an estimate of the risk, and it so happens that if the event in question is
uncommon (say affecting <10% of the population), then the values of RR and OR would be nearly identical. For common events, however, the value of OR can increase remarkably whereas the value of RR stays constrained that is does not increase so much.

To give an interesting example of OR and RR in epidemiological studies, let us turn to the popular Indian street food panipuri. A panipuri man visits a large housing complex with 1000 residents, and 250 of them enjoy his panipuri. The next day, there is an outbreak of dysentery in the housing complex, and the elders call a residents’ meeting. A section of the residents put all blame on the panipuri man and want all such men banned from entering the complex forever. However, not all who develop dysentery have consumed panipuri, and another section of the residents are not so sure and voice their concern that they suspected the main water reservoir to be contaminated and the result was this dysentery outbreak. Let’s put our risk estimates to use to support the decision-making process:

Here we have,

Population = Residents of a housing complex,
Exposure = Consumption of panipuri on day 1,
Outcome = Outbreak of dysentery on day 2.

Counting the exposures and the outcomes, we have our 2 × 2 tables as:

| Group      | Outcome  | Row total |
|------------|----------|-----------|
|            | Present  | Absent    |          |
| Exposed    | 12       | 238       | 250      |
| Nonexposed | 7        | 743       | 750      |
| Column total | 19  | 981       | 1000     |

Thus, the OR now is (12 × 68)/(13 × 7) = 8.97.

While, the RR now is (12 [12 + 13])/7 [7 + 68]) = 5.14.

What has happened? This time, the RR still indicates an over five times increase in risk, but the OR indicates a much larger quantum of risk. Remember what we said earlier. For uncommon events (in the earlier example 0.19% of the residents are affected), the values of RR and OR are nearly the same. However, for common events (in the latter example 19% of the residents are affected), the value of the OR may increase disproportionately while that of RR is constrained.

For meaningful interpretation, the RR or the OR should be presented with the corresponding 95% CIs. If both the lower and upper confidence limits are >1, then definitely the exposure is increasing risk. If both limits are <1, then the exposure is actually protecting against the outcome. However, if the CI includes 1, then the conclusion will not carry so much weight. The calculation of 95% confidence limits for these ratios requires more complicated mathematics than CI for a simple proportion, and we will skip the formulas here.

It is worthwhile to remember that OR is always further away from 1 than the RR for the same set of data. If exposure increases risk, then the relation OR > RR > 1, holds. If exposure reduces risk, the relation that holds is OR < RR < 1. The OR is a symmetric ratio in that the OR for the disease given the risk factor is the same as the OR for the risk factor given the disease. ORs also form part of the output when carrying out logistic regression, an important statistical modeling technique, in which the effects of multiple predictors on a binary outcome variable can be examined simultaneously. Logistic regression provides adjusted OR for individual risk factors that is the OR emerging when the risks from all other factors considered in the model are kept constant.

Both the OR and the RR can also be used in the context of clinical trials to assess the success of the treatment relative to the control or the risk of an adverse event.
compared to control. However, note that using OR carries the potential pitfall of exaggerating risk as we saw earlier. In a clinical trial situation, therefore, it is preferable to state the RR rather than the OR.

**Rate Ratio**

In longitudinal studies, with the possibility of unequal or incomplete follow-up periods for different subjects, it is preferable to use the rate ratio rather than the RR. The rate ratio is analogous to the RR but is calculated as:

\[
\text{Rate ratio} = \frac{\text{Incidence rate in exposed group}}{\text{Incidence rate in nonexposed group}}
\]

Thus, the calculation uses incidence rate rather than incidence.

Consider an often-quoted example from The Nurses’ Health Study. One of the effects studied by this large prospective cohort study was influence of hormone replacement therapy (HRT) on coronary artery disease (CAD) in postmenopausal women. The investigators calculated the incidence rate of CAD in postmenopausal women who had been taking HRT and compared it to the incidence rate in their counterparts who had not taken HRT. The findings are summarized in Table 2.

Therefore, the incidence rate in those using HRT was 30/54308.7 = 55.2/100,000 person-years while the rate in those not using HRT was 60/51477.5 = 116.6/100,000 person-years. This gives a rate ratio of 0.47, suggesting that women who used postmenopausal HRT had about half the rate of CAD compared to women who did not. Although this has also been interpreted as postmenopausal women using HRT had 0.47 times the risk of CAD compared to women not using HRT, it is more precise here to refer to ratio of rates rather than risk.

The applications of the three ratio measures in assessing risk in epidemiological studies have been summarized in Table 3.

Recall that odds of an event is defined as \( P (1 - P) \), where \( P \) denotes the probability of occurrence of the event. The following formula may be used to derive the probability of occurrence of an event, given the odds of the event:

\[
p = \frac{\text{Odds}}{\text{Odds} + 1}
\]

We have seen earlier that using OR carries the potential pitfall of exaggerating risk. However, it is also possible to estimate the RR by adjusting the OR as proposed by Zhang and Yu:

\[
\text{RR} = \frac{\text{OR}}{(\text{P}_{\text{NE}} + 1) + (\text{P}_{\text{NE}} \times \text{OR})}
\]

Where, RR is the relative risk, OR is the odds ratio, and \( \text{P}_{\text{NE}} \) is the proportion of those nonexposed who develop the outcome.

It is evident that as the outcome becomes rare, \( \text{P}_{\text{NE}} \) approaches zero and OR then approximates the RR.

**Financial support and sponsorship**

Nil.

**Conflicts of interest**

There are no conflicts of interest.

**Further Reading**

1. Centers for Disease Control and Prevention (CDC), USA. Lesson 3: Measures of risk. In: Principles of epidemiology in public health practice. 3rd ed. [Online Self-Study Course SS1978/. Available from: https://www.cdc.gov/ophss/csels/dsepd/ss1978/lesson3/section5.html. [Last accessed on 2017 Jan 30].
2. Grimes DA, Schulz KA. Odds and odds ratios. Obstet Gynecol 2008;111:423-6.
3. Tramèr MR, Walder B. Number needed to treat (or harm). World J Surg 2005;29:576-81.
4. Zhang J, Yu KF. What’s the relative risk? A method of correcting the odds ratio in cohort studies of common outcomes. J Am Med Assoc 1998;280:1690-1.
5. Nelson NJ. Nurses’ health study: nurses helping science and themselves. J Natl Cancer Inst 2000;92:597-9.