In epidemiology, a risk factor, or exposure, is an event, condition or characteristic which modifies the risk of an event or outcome. The relationship between exposure and outcome is the effect of the exposure.

**Why Should Oncologists Worry About Risk Factors?**

When a patient has been diagnosed with cancer, the risk factors that caused it might not be of great importance to the oncologist who is treating her. However, it is still important to know about the types of study that investigate risk factors, not least because improved survival and life expectancy of cancer patients have led to an increase in the risk of second cancers (Oeffinger et al, 2013), partly due to treatment effects (Kamran et al, 2016; Morton et al, 2014) and partly due to the risk factors that were responsible for the first cancer (Berrington de Gonzalez et al, 2011). Addressing behavioural risk factors may reduce subsequent risk for the patient (Khuri et al, 2001) and family members may also seek information on reducing their cancer risk (Bottorff et al, 2015; Howell et al, 2013; Radecki Breitkopf et al, 2014). Furthermore, all physicians have a responsibility to give advice that might prevent ill health, and to be aware of the strengths and limitations of the evidence supporting this advice.

**Measurement of Risk**

Risk is defined as the number of events divided by the number of people at risk. When measured over a specified period of time, it is described as the incidence rate. Differences in risk due to an exposure may be expressed as a ratio or a difference.
Risk Number of events/number of people at risk

**Risk ratio**  Risk of exposed/risk of unexposed
- measures the strength of the effect
- is independent of the population risk

**Risk difference**  Risk of exposed - risk of unexposed
- describes the number of additional cases due to the exposure

Excess or **attributable risk** (Parkin, 2011; Whiteman et al, 2015)
- is the difference in the risk of a condition between an exposed population and an unexposed population

**Risk ratio and risk difference**
In a study of hormone replacement therapy (HRT) (Jones et al, 2016), 500 out of 20 114 non-users and 52 out of 1612 users of combined HRT developed breast cancer (Table 1). The risk to users was 3.6% and to non-users 2.5%, giving a risk ratio of 1.30 (i.e. the risk to users was 30% greater). The difference in risk was 0.74%, equivalent to 12 (1612 × 0.74%) additional cases of cancer in the 1612 users.

<table>
<thead>
<tr>
<th></th>
<th>All women</th>
<th>Cases</th>
<th>Risk (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-users</td>
<td>20 114</td>
<td>500</td>
<td>500/20 114=2.5%</td>
</tr>
<tr>
<td>Oestrogen/progestogen HRT</td>
<td>1612</td>
<td>52</td>
<td>52/1612=3.2%</td>
</tr>
<tr>
<td>Risk ratio (3.2%/2.5%)</td>
<td></td>
<td></td>
<td>1.30</td>
</tr>
<tr>
<td>Risk difference (3.2%–2.5%)</td>
<td></td>
<td></td>
<td>0.74%</td>
</tr>
</tbody>
</table>

**Causation**
Risk factor epidemiology tries to separate the effects of the exposure being investigated from all other exposures. This is important because cancer may develop following a series of different exposures over a long period, so the identification of all possible exposures is challenging.
Establishing Causation

Study conditions in epidemiology are difficult to control, so a single study is rarely definitive, and evidence of causation depends on accumulated evidence. Interpretation of this evidence may be controversial.

**Mobile phones and brain cancer**

The INTERPHONE (INTERPHONE Study Group, 2010) and other large studies (Benson et al, 2013) have produced strong evidence that there is no association between mobile phone use and brain cancer, but controversy continues concerning a range of methodological issues (Lagorio and Röösli, 2014; Morgan et al, 2015).

The epidemiologist Bradford Hill (Hill, 1965) proposed certain aspects of a study which suggest causation (Table 2).

**Table 2 Bradford Hill’s Criteria for Causation**

- Strength: An exposure which increases the risk of the outcome by 5% is less convincing than one which doubles it
- Consistency: Has the association been repeatedly observed in different places, circumstances and times?
- Specificity: Is the association limited to particular sites and types of disease?
- Temporality: Does the exposure precede the outcome?
- Biological gradient: Does the association show a dose–response curve?
- Plausibility: Is the causation biologically plausible?
- Coherence: This is related to plausibility – does the effect cohere with the generally known facts of the natural history and biology of the disease?
- Experiment: If some preventive action is taken, does it in fact prevent the outcome?
- Analogy: Has a similar exposure been shown to be associated with a similar outcome?

**Study Design**

Cancer risk factors are often suggested by observing variation in cancer incidence or mortality between populations differentiated by geography, time, occupation or other characteristics. Hypotheses developed from these observations are tested in analytical studies. These are typically cohort or case-control studies, but sometimes a randomised trial (see Chapter 6) might be used.
Types of Epidemiological Study

Table 3  Advantages and Disadvantages of Different Study Types

<table>
<thead>
<tr>
<th>Study type</th>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cohort study</td>
<td>Clear sequence of events</td>
<td>Large numbers of participants needed with long follow-up period, so expensive and often slow</td>
</tr>
<tr>
<td></td>
<td>Risk can be measured</td>
<td>New exposures difficult to add</td>
</tr>
<tr>
<td></td>
<td>Low risk of selection bias</td>
<td>Loss to follow-up</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Change in exposure status during study</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Risk of confounding</td>
</tr>
<tr>
<td>Randomised trial</td>
<td>Clear sequence of events</td>
<td>Large numbers of participants needed with long follow-up period, so expensive and often slow</td>
</tr>
<tr>
<td></td>
<td>Risk can be measured</td>
<td>New exposures difficult to add</td>
</tr>
<tr>
<td></td>
<td>Low risk of bias or confounding</td>
<td>Loss to follow-up</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Change in exposure status during study</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ethical issues</td>
</tr>
<tr>
<td>Case-control study</td>
<td>Relatively small number of participants needed</td>
<td>Risk cannot be calculated</td>
</tr>
<tr>
<td></td>
<td>Disease objectively confirmed</td>
<td>Prone to selection bias, recall bias and confounding</td>
</tr>
<tr>
<td></td>
<td>No follow-up period needed; no drop-outs</td>
<td>Limit to exposures studied</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Difficult to acquire biological samples</td>
</tr>
</tbody>
</table>

Cohort studies

A cohort is a group of people followed over a period, some of whom will have the exposure of interest and some of whom will have the outcome of interest. Participants are assessed for many exposures in addition to that under investigation and often have biological samples taken. For rare exposures, it is necessary to find cohorts with a high prevalence of exposure, such as occupational groups (Kachuri et al, 2016), while general population cohorts are used for more common exposures (Riboli, 2001). A randomised trial can be thought of as a type of cohort study where the exposure is randomly assigned by the researcher. Field trials are the custom in cancer epidemiology, where participants in the community are randomised, either individually or by group (e.g. by area of residence or clinic attended).
Case-control studies

Case-control studies begin with identified cases of cancer whose exposures are compared to those of a group of people without cancer (controls). Both groups are drawn from the same source population. The source population may be patients attending a hospital or clinic, the population of a region or other defined population. The control group is chosen at random from this source population. Sometimes, cases and controls are drawn from an existing cohort. This would be a nested case-control study which provides better quality information on exposures.

Sources of Error in Risk Factor Studies

The errors which occur in studies of causation are of two kinds: systematic and random.

- **Systematic error** is unaffected by study size
- **Random error** decreases with increasing study size

Systematic error

Systematic errors are divided into bias and confounding.

- **Bias** can be considered as an error in the conduct of a study (selection bias, measurement bias)
- **Confounding** is an error in study design or interpretation of study results
Bias

**Selection bias.** Selection bias occurs when the exposed and unexposed populations differ in ways (other than the exposure) which affect the outcome. Selection bias can give rise to the ‘healthy worker’ effect, where the effect of an occupational exposure is countered by the overall better health of those in active work (Zielinski et al, 2009). Selection bias may also occur if participants volunteer for the study for reasons related to the exposure, e.g. interest in a healthy lifestyle.

Bias is difficult to avoid in the selection of the controls for case-control studies. They may be chosen from patients with non-cancer conditions attending the same hospital or from people living in the same area or attending the same family doctor, and so may have risk factors in common with cases.

**Measurement bias.** Exposure measurement: Bias in recall of self-reported exposures is common in case-control studies. Bias may be *differential* between cases and controls, as patients with cancer are more likely to recall a specific exposure, or it may be *non-differential*, due to under-reporting of factors such as alcohol and tobacco intake. Differential bias may lead to over- or under-estimation of the effect, but non-differential bias will always lead to under-estimation. Where possible, self-reported exposures should be independently validated.

Outcome measurement: Bias in outcome measurement is uncommon in cancer epidemiology, although cancer diagnoses may be missed in cohorts for which the *follow-up* is inefficient. Overdiagnosis, or earlier diagnosis, may occur in cohorts where the exposed participants are more intensively monitored.

**Confounding**

Confounding is a common source of error in interpretation. A confounder is something which affects the outcome but not the exposure of interest, and is correlated with the exposure. For instance, heavy drinkers tend to smoke, which means that high alcohol consumption is associated with, but does not cause, lung cancer. Smoking is therefore a confounder of
the relationship between alcohol and lung cancer. Confounding occurs frequently in cancer studies, due to the large number of potential carcinogenic exposures. While bias can be minimised by adherence to good study design and practice, minimising confounding requires a thorough knowledge, measurement and analysis of potential exposures and is usually part of study analysis as well as design.

**Random Error**

The relation between exposure and outcome is unpredictable at the individual level, and measures of effect in individuals will be randomly distributed around some best estimate (e.g. an average). The usual measure for showing the scatter around the estimate is the 95% **confidence interval**. There are various interpretations of this interval, but in practice it is used to test if the data are consistent with some hypothesis (see also Chapter 8). Random error reduces with study size but can also be reduced by study design and conduct and by having a homogeneous study population.

**Statistical Testing**

**Statistical testing** determines how consistent the measured effect is with a hypothesised effect (see Chapter 8). The hypothesis is usually that there is no effect, or that there is no difference between two effects (**null hypothesis**). Conventionally, if the 95% confidence intervals of the measured effect do not overlap those associated with the null hypothesis, it is considered that there is a real effect. Confidence intervals are more informative than probabilities (**p-values**) which give little information about the underlying data.

Risk ratios and **odds ratios** are conventionally presented as unadjusted and adjusted. The **unadjusted ratio** is the simple risk ratio or odds ratio (risk exposed/risk unexposed). On the other hand, an **adjusted ratio** arises from statistical models which allow for the effects of other variables and confounders (e.g. age, sex, smoking, body mass index) which may affect the risk. Table 4 shows an example of unadjusted and adjusted ratios and their confidence intervals.
**Table 4**  Unadjusted and Adjusted Odds Ratios and 95% Confidence Intervals for Colorectal Cancer Risk Associated With Duration of Observed Insulin Exposure

<table>
<thead>
<tr>
<th>Cases</th>
<th>Controls</th>
<th>Unadjusted odds ratio (95% confidence interval)</th>
<th>Adjusted odds ratio (95% confidence interval)*</th>
</tr>
</thead>
<tbody>
<tr>
<td>No insulin therapy (reference)</td>
<td>107 (83.6)</td>
<td>1084 (87.5)</td>
<td>1.0</td>
</tr>
<tr>
<td>≥5 years of insulin use</td>
<td>4 (3.1)</td>
<td>15 (1.2)</td>
<td>2.8 (0.9–8.5)</td>
</tr>
</tbody>
</table>

*Adjusted for sex and 7 other variables.

**Interpretation**

How important is the effect? Two factors determine the clinical importance of an effect:
- The size of the effect
- The frequency of occurrence of the exposure

Large effects, even with wide confidence intervals, should not be ignored if they fulfil criteria of plausibility. Small, *statistically significant* effects are common in large studies, but may be artefactual. However, small effects with high exposure prevalence may have public health importance. Where the background risk is low, risk difference is more informative than risk ratio, because the risk ratio may exaggerate the importance of an effect. The STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) initiative has produced a detailed guide on the reporting and interpretation of observational studies (Vandenbroucke et al, 2007), which describes how these studies should be reported.

**Representativeness**

Studies of cancer risk factors are investigations of aetiology, which are presumed to have a biological basis. Although there may be differences in susceptibility between populations, the effects of risk factors are usually similar in all populations. Good study design is therefore more important (Doll et al, 2004) than the issue of whether the participants are representative of the wider population.
Publication Bias

Many initial studies of risk are small and poorly designed. If they test a novel hypothesis, they are less likely to be published if they fail to support this hypothesis. If published, they are likely to be followed by larger studies, which are more likely to be published. Small negative studies of risk tend to be under-reported, leading to bias in reviews and meta-analysis. Figure 1(a) shows the forest plot of a meta-analysis (see Chapter 9) of the risk of prostate cancer in first-degree relatives of prostate cancer patients (Bruner et al, 2003). Figure 1(b) shows a funnel plot of the same data. The vertical dashed line indicates the weighted average, around which individual studies should be symmetrically grouped. The smaller studies (at the bottom) are skewed to the right, suggesting that smaller negative studies were less likely to be published, causing publication bias.

Figure 1  (a) Relative risks of prostate cancer in men with a history of prostate cancer in a first-degree relative. (b) Funnel plot for first-degree relatives. The circles represent the estimates of the log relative risk for each study and the horizontal lines are 95% confidence intervals.

Conclusions

While the European Code Against Cancer (International Agency for Research on Cancer, 2017) has only 12 proven recommendations for action to reduce risk, a PubMed search for ‘cancer prevention/risk factors’ yields over 130 000 citations. This prompts the question: how, and why, should a busy clinician deal with all this evidence? It is tempting to wait for consensus to be summarised in systematic reviews and meta-analyses (see Chapter 8). However, these vary in quality, may not be up to date and should not be regarded as a substitute for critical reading of key reference papers. Guidelines and checklists help in making an assessment of the evidence, but it is also important to assess the practical importance of the findings. Many ‘positive’ reports turn out to have little practical impact in the real world. It is the responsibility of all cancer clinicians to give cancer prevention advice, but to be aware of the strengths and limitations of the evidence.

Declaration of Interest:
Dr Comber has reported no conflict of interest.

Further Reading


References


