Tools of the Trade:

THE NOTION OF "RISK"

Risk, Causation, Relative Risk, and Odds Ratio Defined

The popular media regularly reports that some activity or substance leads to a particular disease or condition. It would seem that everything that tastes good or is otherwise enjoyable is bad for your health. Most of this research involves an assortment of statistical designs that attempt to quantify the "risk" of some particular outcome associated with one exposure or another.

It is important to understand that risk and cause are different concepts. In a classical scientific context, causation requires the causal factor be both "necessary" and "sufficient" to the outcome. That is, the factor must be present for the outcome to occur, and the factor is all that is required for the outcome to happen. In the biological sciences, cause and effect relationships can be (and usually are) complex with a multitude of factors being interrelated. An adverse outcome might be related to numerous factors and those factors can be linked to multiple outcomes. Therefore, in medicine and health, causation is often inferred under much less rigorous, and often disputed, criteria. Associations that are mathematically unlikely to occur randomly may be used to suggest causation. In public health, the minimal operational requirements to propose that a factor contributes to an outcome is that the presence of the factor increases the probability that the adverse outcome results, and that removal of the factor reduces the probability of an adverse outcome occurring. It must be constantly borne in mind that it is possible that this criteria may be met and yet the factor has no actual bearing on the outcome. Additionally, it is expected that, for statistical associations to suggest causality, other factors that may impact the outcome have been controlled experimentally or statistically and that the relationship between the factor and the outcome is biologically plausible.

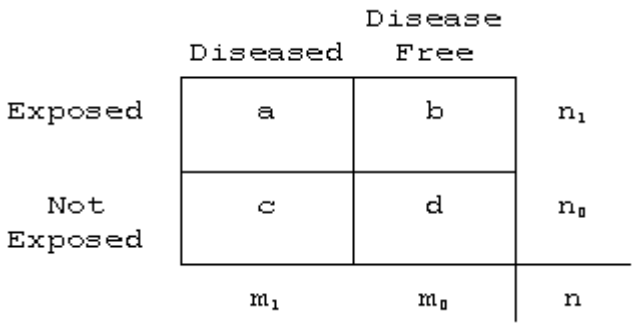
Risk is a term used to quantify the probability of the occurrence of some event given the occurrence of another event. It is an important tool for our understanding of medicine and health but like any tool it must be properly used. Cigarette smokers die from lung cancer much more often than non-smokers, and more people die in hospitals than anywhere else. The statistical risk of dying is high for both of these factors (smoking, hospitals). Nearly all unbiased researchers feel that cigarette smoking causes lung cancer deaths based to a large extent on the statistical evidence. However, few believe that the increased probability of dying in a hospital suggests that hospitals cause death. The increased risk of dying in a hospital is because the best treatment for most life threatening illnesses is only available in hospitals.

Risk is the likelihood of some outcome occurring given the presence of some factor. The outcome is generally referred to as a disease although it might also be death or some adverse medical condition. The factor whose relationship with the outcome we are interested in is generally referred to as the exposure. The exposure could be an environmental/behavioral exposure, a medical treatment or a biological condition. In the health sciences, the measure of risk we are usually interested in is the Relative Risk, although other measures of risk such as an Odds Ratio might be dictated by the nature of the study.

The Relative Risk (RR) is a ratio of the proportion of disease among the exposed relative to the proportion of the disease among the unexposed.

The Odds Ratio (OR) provides an estimate of the relative risk when information on the population at risk (total number of exposed cases) are unavailable. Research using a case control design is the most common use of an odds ratio.

If we prepare a contingency table (2x2), the relative risk (RR) is determined as follows:



$$\mathbf{RR = (a / n_1) / (c / n_0)}$$

This calculation requires that information regarding the exposed population is available. If the marginal values (n_1, n_0, m_1, m_0) are representative of the population under study, either reflecting the actual study cohort or a true random sample, then the relative risk (RR) may be directly calculated.

When information regarding the exposed population is unknown, then the odds ratio (OR) can be used to estimate the relative risk if the disease is rare and the exposure is common. This estimation is used when information is lacking for the exposed population, such as in case control studies where the selection of cases and controls is based on a research design which minimizes the impact of other factors on the analysis. The OR is calculated as follows:

$$\mathbf{OR = ad / bc}$$

When data from real observations are assigned to the cells of the table, a positive value is calculated whose range could fall between zero and infinity. A value of one equals unity or an indication of no risk. A value less than one indicates a protective relationship, and a value of greater than one indicates a deleterious relationship. If a risk (RR) is calculated as 2.34, it may be stated that persons exposed are 2.34 times more likely to contract the disease than those not exposed.

It is important to note that the risk may or may not be statistically significant. That is, a particular value for the observed risk may be simply a random occurrence. Statistical significance represents the probability that the observed risk will occur randomly when no real risk exits. There are a variety of statistical procedures that can be used to identify significance. These tests generally are used to reject the null hypotheses (H0) that there is no significant difference

between the observed and expected measure. An extension of significance testing is the calculation of confidence limits which provides the range of values that the risk measure might assume if the test were replicated. This confidence interval is most commonly calculated at a 95 percent confidence level, although other levels of confidence (99 percent, etc.) can be used. At the 95 percent confidence level, it is expected that the risk measure will fall within the range calculated 19 out of 20 times. When the confidence interval for the risk measure includes the value of 1.0 (or unity), the risk is not considered significant.

For example, if calculation of the 95 percent confidence interval for the RR value of 2.34 were to result in the following range of values:

0.897 - 7.34

The risk would not be considered statistically significant because the value of 1.0 is included in this range or confidence interval.

See Parts I and II of "Exploring Risk Relationships" for real working examples of the determination of risk.