Understanding Research

Risk and Cause in Medical Science

Rashmi Kumar
Sandeep Kumar

The complexity of biological systems makes the question of disease causation also a complex one. Before we declare any agent or factor to be a cause of a disease, we must apply the criteria of causal inference to decide if it is indeed involved in causing the disease in question. The earliest known 'postulates' for causality in medical science, put forward by Koch in 1877(1), are as follows: (*i*) The agent must be shown to be present in every case of the disease by isolation in pure culture; (*ii*) The agent must not be found in patients with other diseases; *(iii)* Once isolated, the agent must be capable of reproducing the disease in experimental animals; and *(iv)* The agent must be recovered from the experimental disease produced.

These postulates were of course true for infectious diseases. However, science has come a long way since then and it is now recognized that multiple factors operate in disease production—both infectious and non-infectious. Thus, while *Mycobacterium tuberculosis* may be responsible for causing tuberculosis, there are other factors such as nutrition and immunological status of the host which determine which of those exposed to the bacillus develop the disease. Also, an 'association' may be found between a disease and a factor, *e.g.,* low roughage diet and breast cancer, but this does not necessarily imply a causal role to the factor. Thus, the original postulates of Koch have given way to the more 'holistic' criteria stated by Hill(2):

1. *Strength of the association.* The stronger the association between the disease and a factor, the stronger is the case for causality. As an exmaple, it may be found that small for gestational age babies are six times as likely to be hypoglycemic than normal ones. This is a reasonably strong association and is likely to represent a true causal relationship.

2. Consistency. Has the observation been made in different places, at different times by different people? The association should be consistently found.

3. *Appropriate time relationship: A* cause must precede an effect. Diseases like cancer usually have a long latent period and if the time period between the putative cause and effect were too small, one should be cautious about accepting this as a causal relationship.

4. *Biological gradient.* The presence of a dose response curve is a powerful argument for an association being causal.

From the Clinical Epidemiology Unit, King George's Medical College, Lucknow 226 003.

Reprint requests : Dr. Rashmi Kumar, Department of Pediatrics, K.G. Medical College, Lucknow 226 003.

The causal nature of birth asphyxia in cerebral palsy is strengthened by the step wise increase in incidence of cerebral palsy with severer degrees of asphyxia.

5. *Biological plausibility.* The causal inference between a factor and a disease should make biological sense. There may be an association between the use of paper in daily life and breast cancer but there is to date no biological plausibility for this to be a causal relationship. Also, what is biologically plausible today may not always hold true.

6. *Specificity.* A specific association implies an almost one to one association between a suspected causual agent and the ensuing disease. However one to one relationships are not frequent. Multicausation is more likely for a disease than a single cause; *e.g.,* cerebral palsy may be caused by other factors as well. Thus lack of specificity does not necessarily rule against causation.

7. Coherence of the evidence. The cause and effect interpretation of our data should not seriously conflict with the generally known facts about the natural history and biology of the disease.

8. *Experiment:* Sometimes experimental findings are available; *e.g.,* the effect may be abolished or diminished by removing a supposed cause. This also is a powerful argument linking cause and effect.

9. *Analogy.* It would be reasonable to ascribe causation if a similar mechanism has been shown to operate for another condition; *e.g.,* our knowledge of the effects of rubella in pregnancy might make us more willing to accept findings relating to other viruses in pregnant women.

As Hill himself pointed out(2), 'None of the above nine criteria can bring indisputable evidence for or against the cause and effect hypothesis and none can be required as a *sine qua non',* but together they form a useful guide to suggest causation in the absence of rigorous experimentation. There are many factors—both known and unknown—operative in disease production. Many factors may contribute to cause one disease and one factor may play a role in the etiology of many diseases. All this has ushered in the concept of 'risk factors'. Thus a risk factor is a factor associated with a higher occurrence of disease. Some risk factors are not causal, but may only be 'indicators'. The exact relationship of a risk factor with the disease in question may not be understood but it is known to "contribute in some way". Thus the expression, in a way is a confession of our ignorance about the mechanism of disease production. Multiple risk factors interplay and have synergistic or antagonistic effects on the production of disease. Risk factors may be racial, *e.g.,* cystic fibrosis in Caucasians, Tay Sachs disease in Jews; genetic, *e.g.,* certain HLA types are known to be associated with ankylosing spondylitis; demographic, *e.g.,* thyroid disorders in females and atherosclerosis in old age; physical, *e.g.,* hypertension is a risk factor for stroke; environmental, *e.g.,* pollution as a risk factor for respiratory infections; behavioral, *e.g.,* smoking and sedentary habits are risk factors for coronary artery disease. The term 'risk' describes the likelihood that people exposed to certain risk factors will acquire a certain disease. Exposure can take place over a single point in time as when a community is exposed to a toxic gas, or it can be over a period

of time, as is usual for most chronic diseases, *e.g.,* sun exposure for melanoma, and cigarette smoking for coronary artery disease. If the factor being studied is not yet established as a risk factor, we call it exposure/study/predictor variable or factor, while the presence or absence of disease is the outcome factor or variable(3,4).

Measures of Risk

One of the commonly used measures of risk is the *'relative risk'* or 'risk ratio' (RR). It tells us how may times more likely exposed persons are to get the disease relative to nonexposed persons and is the ratio of the incidence of disease in exposed to incidence of disease in nonexposed.

$$RR = \frac{\text{Disease in exposed}}{\text{Disease m nonexposed}}$$

If we want to estimate the RR of cerebral palsy in very low birth weight (VLBW) babies, we would study the incidence of cerebral palsy both in infants with normal and VLBW. Suppose, the incidence of cerebral palsy in infants with VLBW (<1500 g) is 70/1000 and in normal birth weight it is 5/1000.

$$\text{Then the RR} = \frac{70/1000}{5/1000} = 14$$

The relative risk does not tell us much about the magnitude of absolute risk. The relative risk may be quite large and yet the absolute risk will be small if the exposure is uncommon(4).

Attributable Risk (AR).

Another measure of risk is the AR or risk difference. It tells us what is the additional risk of disease following exposure, over and above that in people not exposed. The additional incidence of disease related to that exposure, taking into account the background incidence of disease due to other causes is the AR. It is the incidence of disease in exposed minus the incidence of disease not exposed. In the example above, AR = 70/1000 - 5/1000 = 65/1000; *i.e.,* 65 cases of cerebral palsy per 1000 VLBW babies. If we imagine a population of 1000 VLBW babies, 70 of these would have cerebral palsy. Of these 65 could be attributable to VLBW directly.

Population Attributable Risk (PAR).

This measure tells us how much a risk factor contributes to the. overall rate of disease in a population. Some risk factors may be very strong (in terms of RR) but may be present only rarely and therefore, still not contribute strongly to the disease. PAR is a measure of the excess incidence of a disease in a population, that is associated with a risk factor and is the product of AR and prevalence of that risk factor in the population. Thus, if the prevalence of VLBW in the population is 5%, then

PAR of cerebral palsy due to VLBW = 65/1000 x 5/100
= 325/100,000

Thus in a population of 100,000, 325 would have cerebral palsy attributable to VLBW only.

Estimating Risk

Different research strategies(5) are followed to determine risk and RR. The best way to study these are through experimental studies. Subjects are divided randomly so that they are similar in every way except for the exposure in ques-

tion. This approach is easiest to follow in drug or therapy trials, but may be impossible for many research questions about causation as it may not be possible to manipulate risk factors. Observational cohort studies(5) are also suitable for measuring risk. A population with and without the exposure is followed up in time and the incidence of disease in both groups is studied. We thus have four groups of individuals:

(i) those with exposure and disease - a

(ii) those with exposure but without disease - b

(iii) those without exposure but with disease - c

(iv) those without exposure and without disease - d

The relative risk can be calculated from the 2x2 table *(Table I).*

$$RR = \frac{\text{Disease in exposed}}{\text{Disease in unexposed}} = a/a+b \div c/c+d$$

AR = a/a+b - c/c+d

PAR = AR x Proportion of population exposed.

For example, if we followed up 250 newborns with VLBW and 500 newborns with normal birth weight and we find that 18 of the VLBW babies developed features of cerebral palsy while only 3 of the normal weight babies developed these. The 2x2 table is as shown in *Table II.*

Then the RR = 18/250 ÷ 3/500 = 12

AR = 18/250 - 3/500 = 33/500 = 66/1000

PAR (if the incidence of VLBW is 2%) = 66/1000 x 2/100 = 132/100,000

TABLE I-*The 2 X 2 table*

	Disease +	Disease -
Exposure +	a	b
Exposure -	c	d

TABLE II- *The 2x2 Table for Cerebral Palsy and VLBW*

	Cerebral palsy +	Cerebral palsy -
VLBW +	18	232
VLBW -	3	497

A 95% confidence interval of the RR is calculated to provide a statistic which can be used for comparative studies in two population groups. Multiple risk factors interacting with each other may produce 'confounding'. These can be handled by multivariate analysis.

Cohort studies are often not feasible because of the time, effort and expense involved and especially when the outcome is rare. The measure of association from case control studies, *i.e.,* the *odds ratio* (OR) or cross product ratio approximates the RR provided the incidence of the disease in the population is low. In case control studies, cases of disease and controls are compared with respect to a particular exposure(5). Thus we would take cases of cerebral palsy and matched controls without cerebral palsy and obtain details of birth weight.

In the example above,

OR = ad/bc = 18 x 497 ÷ 3 x 232 = 13.04

Case control studies are often more practical and can be done efficiently especially for rare diseases. However, the

case control design is disreputed on account of various 'biases' which may occur.

REFERENCES

1. Evans AS. Causation and disease: The Henle-Koch postulates revisited. Yale J Biol Med 1976, 49:175-195.
2. Hill AB. Principles of Medical Statistics. London, Lancet, 1971, pp 312-320.
3. Last JM. *A* Dictionary of Epidemiology. New York, Oxford University Press, 1988, pp 5-40.
4. Fletcher RH, Fletcher SW, Wagner EH. Clinical Epidemiology—The Essentials, 2nd edn. Baltimore, Williams and Wilkins, 1988, pp 91-105.
5. Kumar R, Kumar S. Research strategies for the clinician. Indian Pediatr 1993, 30; 1041-1048.