Seminar ¹

Risk and prognosis ²

A. Hofman

Department of Epidemiology and Biostatistics, Erasmus University Rotterdam, PO Box 1738, 3000 DR Rotterdam, Netherlands

Keywords: Clinical epidemiology; Risk estimation; Risk and prognosis

1. Introduction

Epidemiology is concerned with the study of the frequency of disease occurrence. Clinical epidemiology is that part of epidemiology that deals with questions concerning clinical practice. Risk and prognosis are important concepts in clinical research and patient management. Risk of disease can be viewed as the chance, or the probability, that disease occurs. In order to know the probability of disease or death, one has to know the frequency of occurrence of disease. For the study of risk and prognosis the frequency of new disease events (incidence), and not the frequency of existing disease (prevalence) is of importance. This is because it is possible to estimate the probability of disease and thereby the risk of disease by estimating its incidence. The risk of disease can be measured in two ways: directly from the cumulative incidence, with or without use of the life table, and indirectly from the incidence rate.

2. Risk estimation

The most important measure of disease frequency is the cumulative incidence. The cumulative incidence is the number of new disease events in a certain specified time period, divided by the total number of people without the disease (and therefore "at risk") at the beginning of that specified period. Schematically the cumulative incidence (CI) in the period $T_0$ to $T_1$ can be given as follows:

$$ CI = \frac{\text{new cases in period } T_1 - T_0}{\text{number of people without disease at } T_0} $$

The cumulative incidence is a proportion or fraction (e.g., a percentage). It is a probability that directly measures the disease risk, and it can vary from 0 to 1. In clinical medicine one preferably uses the cumulative incidence as a measure of disease frequency, but sometimes one is forced to use another measure: the incidence rate (IR). This also is a measure that uses the number of
new cases, but now divided by the number of
person-years of follow-up:

\[
IR = \frac{\text{new cases}}{\text{number of person-years of follow-up}}
\]

The number of person-years is the product of the
number of persons and the period of observation
of these persons. One hundred persons observed
for 10 years amount to 1000 person-years. The
incidence rate can vary from 0 to + infinity.

The incidence rate is a more complicated mea-
sure of frequency than the cumulative incidence.
It has no simple interpretation, and it is not a
proportion or percentage and it does not directly
measure the probability or risk of disease. How-
ever, it is possible to calculate the cumulative
incidence from the incidence rate. With a reason-
ably low incidence rate and a relatively short
follow-up period the following rule of thumb may
be used:

\[
CI_{\Delta t} = IR \times \Delta t.
\]

As an example, one can calculate the 5-year
cumulative incidence of dementia for a 70-year-
old Dutch man from the incidence rate of demen-
tia in the age category 70–74 years (1.98/1000
man-years). This gives an estimate of the risk to
become demented within 5 years of CI\(_{5}\) = 0.00198
\times 5 = 1%. A more extensive discussion with more
general formulae to estimate the risk from an
incidence rate can be found elsewhere [1].

3. Types of risk

It is possible to distinguish between 3 types of
risk: absolute risk, relative risk and attributable
risk.

The absolute risk is the probability of disease,
complication or death. An absolute risk must be
given for a specified time period. This can be a
concretely defined time period, like the 10-year
risk of a myocardial infarction in a 40-year-old
man, but also a more abstract time period, like
the life-time risk of leukaemia, which denotes the
probability that someone will suffer from
leukaemia in his lifetime. In clinical medicine,
absolute risks are used in various forms. The 5-
or 10-year survival is a cumulative incidence. The
lethality or case fatality is generally used as the
risk of death in patients who suffer from a certain
disease; it is a cumulative incidence of death. The
term attack rate is often used for infectious dis-
esases and it is a cumulative incidence over a short
period of time. In clinical practice it is important
to know the absolute risk as specifically as possi-
ble. It is of little value to know the probability of
a myocardial infarction in general. One would
like to know that probability for men and women,
for different ages, for persons with or without
hypertension and with or without a positive fam-
ily history for a myocardial infarction. In other
words, one would like to know the risk of disease
specified according to determinants of disease.
These determinants are often called risk indica-
tors because the disease risk varies with categ-
ories of the determinants. A risk indicator with a
causal relation with the disease is referred to as a
risk factor. Age and sex are risk indicators for
myocardial infarction, whereas blood pressure and
serum lipids are risk factors; age is a risk indica-
tor for lung cancer, and smoking habits a risk
factor.

A measure of risk that is often used is the
relative risk. This measure has two forms: the risk
ratio (RR) and the risk difference (RD). The term
“risk ratio” is sometimes used as a synonym for
the more general term “relative risk”. It specifies
the number of times that the absolute risk is
higher in one category of the determinant (the
index group) than in another category (the refer-
ce group). As an example may serve that the
risk ratio of smoking and lung cancer is about 10.
This means that the absolute risk of lung cancer
among smokers (\(R_1\)) is 10 times higher than the
risk of lung cancer among non-smokers (\(R_0\)). The
risk ratio can be calculated by dividing the abso-
lute risks of the index group (\(R_1\)) and the refer-
ce group (\(R_0\)):

\[
RR = \frac{R_1}{R_0}.
\]

The risk ratio is a measure of risk with the lowest
value of 0 and the highest value of + infinity.
The null-value, that is the value in which the
absolute risks in the index and reference groups are the same, is one.

The risk difference is the measure that gives the difference between two absolute risks. If the risk in the index group is given by \( R_1 \) and the risk in the reference group by \( R_0 \) then the risk difference is given by:

\[
RD = R_1 - R_0
\]

This measure can be used to calculate how much the absolute risk of a patient is higher than that of a reference patient. The reciprocal of the risk difference, i.e. \( 1/RD \), can be used to calculate the number of patients that have to be treated to prevent one “event” (complication, death). An example may be taken from a trial of patients with mild hypertension, conducted by the British Medical Research Council. In this MRC trial, patients with mild hypertension were treated with either an antihypertensive drug (beta-blocker or diuretic) or with a placebo. In comparing the efficacy of the antihypertensive treatment with the placebo treatment it was observed that the group on drug treatment (the index group) had a 10-year cumulative incidence of cerebrovascular stroke of 1.4% \( (R_1 = 0.014) \), compared to 2.6% \( (R_0 = 0.026) \) in the placebo group. The risk difference was therefore 1.2%. This means that one has to treat for a period of 10 years \( 1/0.012 = 83 \) patients with mild hypertension to prevent one cerebrovascular accident. The same study showed that one has to treat \( 333 \) patients with mild hypertension for 10 years to prevent one myocardial infarction.

A measure of risk which is often used in public health, and less in clinical practice, is the attributable risk (AR). This measure describes the proportion of sufferers from a disease that can be ascribed to one particular determinant or risk factor. The attributable risk can be calculated by dividing the risk difference by the absolute risk in the index group:

\[
AR = \frac{R_1 - R_0}{R_1}
\]

This measure is also referred to as aetiological fraction, because it denotes what part of the risk \( (R_1) \) is the net effect of a certain risk of a certain factor \( (R_1 - R_0) \). A more detailed discussion of the relation between relative and attributable risk can be found elsewhere [1].

**References**