AMERICAN

Journal of Epidemiology

Formerly AMERICAN JOURNAL OF HYGIENE

VOL. 102 OCTOBER, 1975 NO. 4

Reviews and Commentary

DEFINITION OF RATES: SOME REMARKS ON THEIR USE AND MISUSE

REGINA C. ELANDT-JOHNSON

In almost every scientific discipline there is a certain amount of lack of precision and ambiguity in terminology which often upsets and confuses beginners before they become accustomed to the situation and accept it (some never do).

One reason for this might be that authorities in one field borrow terminology from another field in which they are not specialists. Another reason could be a matter of semantics; a particular language may have more than one meaning for the same word; or vice versa, people use two (or more) words as synonyms though, in fact, their meanings are distinct.

The use of the word *rate* in epidemiology, demography, medicine and even in actuarial work suffers from these disadvantages; it is borrowed from physics and biochemistry and misinterpreted; it has more than one meaning in the English language; and the most common error—it is used interchangeably with the term *proportion*, because both are incorrectly assumed to be synonyms of *ratio*.

Before you proceed to the next section, please, take a piece of paper and a pencil and write down your definitions of *ratio*, *proportion* and *rate*. If you have difficulties, or are curious about what has been said about these terms in respectable books in your discipline, compare your definitions with those given by others. Finally, read the rest of this article and argue with me if you disagree with my definitions.

Ratios, proportions, rates

Ratios

Ratio is, in a very broad sense, the result of dividing one quantity by another ($R = a/b$).

In sciences, however, it is mostly used in a more specific sense, that is, when the numerator and the denominator are two separate and distinct quantities; neither is included in the other. Often the quantities are measured in the same units, but this is not essential. For example,

$$\textit{Sex ratio} = (\textit{No. of males})/(\textit{No. of females})$$

$$\textit{Fetal death ratio} = (\textit{No. of fetal deaths})/(\textit{No. of live births}),$$

in a given population.

An *index*, a sort of comparative summary measure of two (or more) phenomena, is often expressed as a ratio. For example,

$$\textit{Weight-height index} = kg/(cm - 100)$$

This work was supported by US National Heart and Lung Institute Contract NIH-NHLI-71-2243 from the National Institutes of Health.

Department of Biostatistics, University of North Carolina, Chapel Hill, NC 27514.

is a ratio and it is used as a measure of obesity.

Proportions

Proportion is a type of ratio in which the numerator is included in the denominator [$p = a/(a + b)$]. For example,

[*No. of males*]/[(*No. of males*) + (*No. of females*)]

in a given community is the *proportion* of males in this community.

Epidemiologists calculate the *proportion of fetal deaths* = (No. of fetal deaths)/(No. of conceptions) and call it (incorrectly) the "fetal death rate." It is a *relative frequency* of fetal deaths among all conceptions and can be used as an estimate of the *probability* of this event.

Generally, the numerator and the denominator in $a/(a + b)$ do not need to be integers. They can be measurable quantities such as, for example, weight, length, space, volume, etc.; in such cases proportions are often also called *fractions*. For example, a mass of one part of a body can be expressed as a fraction of the total mass of the body. In random phenomena, such fractions could be used in estimating probabilities.

Concepts of instantaneous and average rates

Ratios and proportions are useful summary measures of phenomena which have occurred under certain conditions. In particular, in studies of populations, the conditions are often determined by factors such as race, sex, space, and often in a definite period of time (e.g., in a year).

On the other hand, the concept of *rate* is associated with the *rapidity of change* of phenomena such as: chemical reactions (gain or loss of mass, increase or decrease in concentration), birth, growth, death, spread of infection, etc. *per unit of time* or other variable (e.g., temperature or pressure).

Generally, a phenomenon may be described by a continuous function y of another (independent) variable x, i.e., $y = y(x)$.

Rate may be defined as a *measure of change* in one quantity (y) *per unit* of another quantity x on which y depends. Thus if $y = y(x)$ and $\Delta y = y(x + \Delta x) - y(x)$, then the *average rate of change* (i.e., the average change in y per unit of x in the interval $(x, x + \Delta x)$) is $\dfrac{\Delta y}{\Delta x}$.

Since x is usually time and y describes a continuous process in time, $\dfrac{\Delta y}{\Delta x}$ is the *average velocity* of the process. It has sign + or −, depending on whether y increases or decreases with time.

In many situations $\dfrac{\Delta y}{\Delta x}$ varies with Δx. Thus the "true" rate per unit time at the instantaneous time point x is

$$\lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} = \frac{dy}{dx} = \alpha(x). \qquad (1)$$

More precisely, equation 1 is the (absolute) *instantaneous rate of change in y per unit time at the time point x*. It is the *true velocity* of a process (or reaction) at time x. The "velocity curve", $\alpha(x)$, describes the pattern and direction of changes in $y(x)$.

We shall take most of our examples from life tables because they are closely related to the study of epidemiology of chronic diseases.

The basic function of the life table is the *survival function* l_x. It can be considered as a continuous monotonically decreasing function of x. The function $-\dfrac{dl_x}{dx}$, called the *curve of deaths*, describes the speed (*velocity*) of dying.

Relative rates

However, in most chemical and biologic processes not the absolute change in substance per unit of time, but the relative change per unit of time *and* per unit of substance available at this time is more informative.

For convenience, suppose that $x > 0$ is *time*, and $y(x)$ is a "mass" which is exposed to reaction (e.g., decay) at time x.

The (relative) *instantaneous rate of change per unit of mass $y(x)$ and per time unit at the time point x* is

$$\beta(x) = \frac{1}{y(x)} \frac{dy}{dx}. \qquad (2)$$

Since this kind of rate is the most commonly used, the word "relative" is often omitted unless ambiguity might arise thereby.

In chemistry, equation 2 is often called *reaction velocity*. We notice that, in fact, equation 2 can be written as

$$\frac{d \log y(x)}{dx} = \frac{1}{y(x)} \frac{dy}{dx} = \beta(x). \qquad (3)$$

If we know $\beta(x)$, we can evaluate $y(x)$. Integrating equation 3, we obtain

$$y(x) = \exp[\int_0^x \beta(u)du]. \qquad (4)$$

Exponential growth model is an example. Let $P(t)$ denote the population size at time t and assume that the change in population size, $dP(t)$, is proportional to its actual size and to change in time, dt, that is

$$dP(t) = aP(t)dt,$$

where a is a growth rate. Hence

$$P(t) = P(0)e^{at} \qquad (5)$$

Another example is the *force of mortality* in a life table, defined as

$$\mu_x = -\frac{1}{l_x}\frac{dl_x}{dx}. \qquad (6)$$

This is the (relative instantaneous rate of change in survivorship (i.e., rate of dying) of a cohort experiencing a particular mortality pattern as described by the l_x column of a life table.

It is also possible to interpret the rate $\beta(x)$ defined in equation 2 as

$$\beta(x) = \frac{1}{y(x)}\frac{dy}{dx} = \lim_{\Delta x \to 0} \frac{\Delta y}{\int_x^{x+\Delta x} y(u)du} = \lim_{\Delta x \to 0} \frac{\Delta y}{y(x')\Delta x}, \qquad (7)$$

for some x' in $(x, x + \Delta x)$.

Here the integral $\int_x^{x+\Delta x} y(u)du = y(x')\Delta x$ (shaded area in figure 1) may be interpreted as an amount of *mass-time* ("mass $\times$ time") available between x and $x+\Delta x$. We may then define a (relative) *average rate of change* of mass $y(x)$ *in* $(x, x + \Delta x)$ *per unit of time* x, *as*

$$b(x) = \frac{\Delta y}{\int_x^{x+\Delta x} y(u)du} = \frac{\Delta y}{y(x')\Delta x}, \qquad (8)$$

for some x' in $(x, x + \Delta x)$.

In fact, if the mathematical form of $y(x)$ is not known, only $b(x)$ defined in equation 8 can be calculated from the data. This is why we need to use an "average rate" rather than an instantaneous one. It should be noticed that

$$p(x) = \frac{\Delta y}{y(x)} \qquad (9)$$

represents that *fraction* (proportion) of the mass $y(x)$ available at time x, which has been changed (e.g., decayed) in a definite period of time x to $x + \Delta x$. This is not a rate. The rate, defined in (8), might be approximately evaluated as

$$b(x) \doteq \frac{\Delta y}{y(x) + \frac{1}{2}\Delta y}, \qquad (10)$$

where the denominator represents the approximate amount of "mass-time" (Note: Δy is a negative quantity if the mass undergoes decay.)

In life tables, the *central death rate* is defined as

$$m_x = \frac{l_x - l_{x+1}}{\int_0^1 l_{x+t}dt} = \frac{d_x}{L_x}. \qquad (11)$$

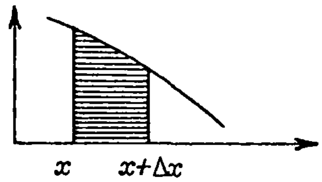

FIGURE 1.

This is a kind of "average rate." Here L_x is the total number of *person-years* lived by the population l_x in the year x to $x + 1$. We recall that in actuarial work, L_x is often approximated by $l_x - \frac{1}{2}d_x$.

The central death rate is estimated by

Age specific death rate

$$= \frac{\textit{No. of observed deaths in age } x \textit{ to } x + 1}{\textit{Population at risk of this age}}.$$

Population at risk, also called in actuarial science "*central* exposed to risk," is interpreted as the equivalent number of persons, each exposed to risk of dying for a full year. In experimental populations, this number is estimated using various approximations. In calendar-year type population data, the *midyear* population size of age x last birthday is used as an approximation to the "true" size of population at risk (i.e., "central exposed to risk").

It seems that it would be useful to clarify the definition of the important life table function

$$q_x = \frac{l_x - l_{x+1}}{l_x} = \frac{d_x}{l_x}, \qquad (12)$$

which is closely related to m_x. It is sometimes called the "*mortality rate*." Clearly, it is not a rate; it is the probability (proportion) of deaths *in a year* (fixed time) but *not per year*, of those who survive to age x; the common arbitrary, though conveniently chosen time unit—one year—probably causes the confusion in using the term "rate" instead of "proportion." Also "population at risk" might be misinterpreted. The number of individuals alive at the *beginning* of the interval is sometimes called by actuaries "initial exposed to risk". Often this distinction is not observed.

Perhaps, a numerical example will make the difference between proportion and rate clearer.

Suppose that a group of 100 individuals (e.g., mice) were alive at the *beginning* of an interval of length, one year, and 40 died during this year. Then the proportion of deaths (i.e., probability of dying in this interval) is 40/100 = 0.40. Suppose that the deaths were uniformly distributed over the year. This implies that the average time of death is the midpoint of the year. Thus during the interval of one year each survivor contributed one full year to the "exposed to risk," while each individual who died contributed, on the average, only ½ of the year. The total number of "exposed to risk" is the total number of "person-years," and in our example, this is $60 \cdot 1 + 40 \cdot \frac{1}{2} = 80$. The central death rate (i.e., the average "rapidity of dying") is 40/80 = 0.50, and this is a different number from 0.40.

Incidence and prevalence

Two rather important epidemiologic terms—*incidence* and *prevalence* "rates"—may also give rise to some ambiguity. Unfortunately, there are no precise definitions of these terms.

For example, the number of *new* cases of a disease which occurs per year in a given community is the *absolute incidence rate;* it does not refer to the population size (Note: the number of deaths per year might be called the "absolute death incidence rate," but it is not customary to use this expression.)

To calculate relative rates, however, we must evaluate the "person-periods" (often person-years) or obtain an estimate of them. If the period is a year, the size of the *midyear* population is used as an estimate of person-years exposure and the relative *incidence rate* is the number of new cases per person per year (or per 1,000 persons per year).

If, however, we calculate the ratio of the number of new cases of a disease to the number of individuals free of the disease at the beginning of the time interval, then this will be a proportion, not a rate. For epidemics, when the development of the disease is very rapid, these two quantities may differ considerably. The proportion can still be called "incidence" if one wishes (perhaps "probability of incidence" would be better), but not a rate.

In contrast, *prevalence*, which is the ratio of the number of cases at a given time to the size of the population at that time is clearly always a proportion. The term prevalence is quite legitimate, but "prevalence rate" is an impossible concept.

Conclusion

In summary, we may say that in sciences, ratios, proportions and rates are precisely defined and cannot be used as synonyms. Ratios are used as indices, proportions are relative frequencies or fractions and often estimate (or are) probabilities of certain events, while rates describe the velocity and direction (patterns) of changes in dynamic processes.

After writing this article, I gave some thought to the question: what *proportion* of research workers in disciplines mentioned at the beginning of this paper will be "converted" and use these definitions, and what might be the "*rate of conversion*". I guess, both will probably be rather low since training and long-established habits may have stronger effect than the "catalytic" power of mathematical arguments.

But *at any rate* (!), I have tried my best to clarify the distinction between ratios, proportions and rates.

Editor's note.—We suspect that most epidemiologists occasionally or regularly misuse one or another of the terms so carefully defined by Dr. Elandt-Johnson. We share her pessimism regarding the likelihood of changing this situation.

We formerly took care to refer only to prevalence *ratios* (never rates—well, hardly ever), but even this expression she indicates is incorrect, the proper term being just prevalence or perhaps prevalence proportions. Alas, so general is the term prevalence rate that many who are aware of this distinction nevertheless use it. Thus, Sir Austin Bradford Hill in the seventh edition of his widely-used and excellent textbook, *Principles of Medical Statistics*, does so.