Table 4 – Duration, mortality (i.e., risk) and force of mortality (i.e., rate) for cholera and phthisis. Source: (Farr, Part II).

Disease	Mean duration (in days)	Mortality (% of all the sick)	Force of mortality (= Mortality rate per 100 sick a year)
Cholera	7	46	2415
Phthisis	730	90–100	50

was almost half of that over 7 weeks based on households (Table 2) and differences were less important.

2.4. Risks and rates

It has taken about 150 years to sort out the properties of risks and rates, clarify their interpretation and produce a theory of their mathematical relationships. We will review here three episodes of this process.

2.4.1. Burden of life destruction and force of mortality

As Superintendent of the General Register Office, England's center for vital statistics, William Farr (1807–1883) was responsible for collecting and reporting information on causes of death (Susser and Adelstein, 1975). In the pamphlet entitled "*On Prognosis*", reproduced *in extenso* in this book (Farr, Part II), Farr illustrates the need for different types of measure of disease occurrence by contrasting an acute infectious disease, cholera, with a chronic infectious disease, phthisis (i.e., tuberculosis). He invokes the following paradox:

> *"Cholera destroys in a week more than phthisis consumes in a year. Phthisis is more dangerous than cholera; but cholera, probably, excites the greatest terror."* (Farr, Part II).

Table 4 shows that almost every tuberculosis patient will die from the disease. The case fatality risk of phthisis is 90–100%. Cholera kills only one of two persons who are affected: its case fatality risk is 46.2%.

Half of the people who get cholera but almost none of those with phthisis will survive. Between cholera and phthisis, it would seem reasonable to prefer cholera, but people fear cholera more than tuberculosis. Why is it so? Farr notes that mortality is

insufficient to characterize the "form and nature of diseases". We need two different measures of disease occurrence:

> *"Diseases may be examined (1) in their tendency to destroy life, expressed by the deaths out of a given number of cases; and (2) in their mean relative 'force of mortality', expressed by the deaths out of a given number sick at a given time."* (Farr, Part II).

Let us consider each of these two ways of examining a disease. For the first parameter, "the tendency to destroy life", Farr gives as synonyms the "probability of death", "mortality" and "death percent". If 990 patients died out of 2,142 cases of cholera, "mortality" is 46.2%. Farr does not use the word "risk", but risk is the term that we would commonly use today. More specifically, this is a "case fatality risk". It expresses the probability that patients with cholera will *die* from their disease. Deaths are in the numerator and sick people are in the denominator.

The second parameter, "force of mortality", is the "quantity eliminated daily by death out of a given constant quantity (e.g., 100) sick". Farr also refers to it as the "mean *rate* of dying per unit of sick time". To compute the force of mortality, Farr divides the number of deaths by the product of the number of persons sick and the average duration during which they were sick. If 2,142 cases of cholera have been sick an average of 7 days each, this corresponds to a total of [7 × 2,142 =] 14,994 days of sickness, or sick person-days. Sick-person days divided by 365 days in a year gives 41 years of sickness or 41 sick person-years. Thus, if 990 die out of 41 sick person-years of cholera, the "force of mortality" is [(990 ÷ 41) × 100) =] 2,415 per 100 sick person-years. The modern synonym of "force of mortality" is mortality rate, and in this example specifically, it is a "case fatality rate". It is the proportion of the cases that will die from their disease *per unit of time*: 2,415 per 100 patients per *year* or 6.6 per 100 patients per *day*.

Distinguishing these two measures of death occurrence allows Farr to explain the paradoxical terror generated by cholera. The data are shown in Table 4. Almost all patients died from tuberculosis (mortality risk = 90–100%), but the death rate is small (50 per 100 per year) and the average duration of the disease is long (2 years). Tuberculosis kills slowly. On the other hand, less than half of the sick will die from cholera (mortality risk = 46%), but the death rate is huge (2,415 per 100 per year) and the average duration of the disease is short (7 days). Cholera appears abruptly, kills rapidly and disappears. Viewed as such, cholera is more frightful.

Why did Farr use the word "force" to characterize a rate? We can speculate that this is in relation to the concept of physical force. Farr must have been familiar with the concept of force defined by the physicist Isaac Newton (1643–1727) in his "*Principia*" (Newton, 1687):

> *"An impressed force is an action exerted upon a body, in order to change its state, either of rest, or of moving uniformly forward in a right line. This force con-*

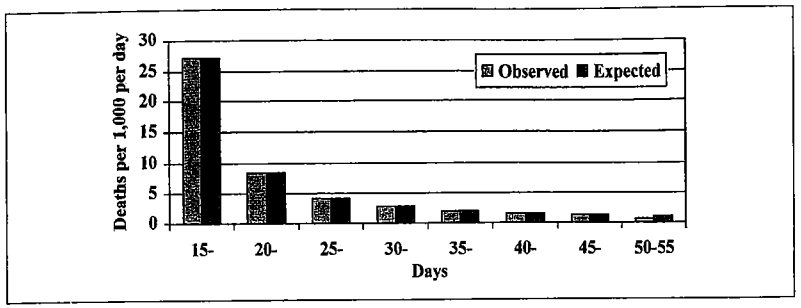

Figure 1
Evolution of the observed and expected death rates from smallpox. Source: William Farr, On Prognosis (Farr, Part II).

sists in the action only; and remains no longer in the body when the action is over." (Cited by Einstein and Imfeld, 1966, p. 11).

A force can be represented by a vector, which has a direction and a velocity. The velocity, that is, the distance covered per unit of time, is by definition a rate. The force of mortality, like a vector, has a velocity and a direction. Mortality rates can go up or down.

Farr notes that predicting the direction in which risk will evolve is crucial for prognosis. The sign of the force indicates whether the rate increases or decreases over time. Indeed, Farr gives the data needed to compute the force of mortality on the 18th, 19th day, etc. of duration of smallpox (Gerstman, Part II). Using the word "*rate*", Farr notes that:

"The rate of mortality [from smallpox] increased from the 5–10 days to 10–15 when it attained a maximum (31.18); it decreased in a determined progression from the next period (15–20 days) to the end." (Farr, Part II).

Farr was mostly interested in the declining part of the rate curve (see Figure 1), which demonstrated some mathematical regularity:

"The decrease begins to take place in geometrical progression; but the tendency to decrease is met by another force that neutralizes part of its effect."
(Farr, Part II).

Again, the use of the force of mortality had a very important clinical implication. In the case of cholera, early treatment was essential because half of the deaths happened in the first 24 hours:

"What the practitioner does he should do quickly."
(Farr, Part II).

2.4.2. The fallacy resulting from neglect of the period of exposure to risk

We speak of a 5-year-risk or a 10-year risk. Whether the risk is over 5 or 10 years is critical for its interpretation. Neglecting the period of exposure to risk can also lead to invalid interpretation of a study result. The British epidemiologist Austin Bradford Hill (1897–1991) described the potential fallacy resulting from neglect of the period of exposure to risk in his textbook "*Introduction to medical statistics*" (Hill, 1939). As it is difficult to write more clearly than Hill, I will quote him here extensively.

"Suppose on January 1st 1936 there are 5,000 persons under observation, none of whom are inoculated; that 300 are inoculated on April 1st, a further 600 on July 1st, and another 100 on October 1st. At the end of the year there are, therefore, 1,000 inoculated persons and 4,000 still uninoculated. During the year there were registered 110 attacks amongst the inoculated persons and 890 amongst the uninoculated. If the ratio of recorded attacks to the population at the end of the year is taken, then we have rates of 110 ÷ 1,000 = 11.0 per cent amongst the inoculated and 890 ÷ 4,000 = 22.3 per cent amongst the uninoculated, a result apparently very favorable to inoculation. This result, however, must be reached even if inoculation is completely valueless, for no account has been taken of the unequal lengths of time over which the two groups were exposed. None of the 1,000 persons in the inoculated group were exposed to risk for the whole of the year but only for some fraction of it; for a proportion of the year they belong to the uninoculated group and must be counted in that group for an appropriate length of time.

The calculation should be as follows:

All 5,000 persons were uninoculated during the first quarter of the year and therefore contribute (5,000 × 1/4) years of exposure to that group. During the second quarter 4,700 persons belonged to this group – i.e., 5,000 less the 300 who were inoculated on April 1st – and they contribute (4,700 × 1/4) years of exposure to the uninoculated group. During the third quarter 4,100 persons belonged to this group – i.e.. 4,700 less the 600 who were inoculated on July 1st – and they contribute (4,100 × 1/4) years of exposure. Finally in the last quarter of the year there were 4,000 uninoculated persons – i.e., 4,100 less the 100 on October 1st – and they contribute (4,000 × 1/4) years of exposure. The "person-years" of exposure in the uninoculated group were therefore (5,000 × 1/4) + (4,700 × 1/4) + (4,100 × 1/4) + (4,000 × 1/4) = 4,450, and the attack-rate was 890 ÷ 4,450 = 20 per cent. – i.e., the equivalent of 20 attacks per 100 persons per annum. Similarly the person-years of exposure in the inoculated group are (0 × 1/4) + (300 × 1/4) + (900 × 1/4) +

Table 5 – Hypothetical example illustrating the fallacy resulting from neglect of the period of exposure to risk. Source: Table XVII, in (Hill, 1939, p. 130).

Inoculated at each point of time	Inoculated		Uninoculated	
	Exposed to risk in each quarter of the year [A]	Attacks at 5 per cent per quarter [B = A × 0.05]	Exposed to risk in each quarter of the year [C]	Attacks at 5 per cent per quarter [D = C × 0.05]
Jan. 1st, 0	0	0	5,000	250
April 1st, 300	300	15	4,700	235
July 1st, 600	900	45	4,100	205
Oct. 1st, 100	1,000	50	4,000	200
Total at end of the year	1,000	110	4,000	890

(1,000 × 1/4) = 550, for there were no persons in this group during the first three months of the year, 300 persons during the second quarter of the year, 900 during the third quarter, and 1,000 during the last quarter. The attack-rate was, therefore, 110 ÷ 550 = 20 per cent, and the inoculated and uninoculated have identical attack-rates. Neglect of the durations of exposure to risk must lead to fallacious results and must favor the inoculated. The figures are given in tabulated form (Table XVII).

Fallacious Comparison – Ratio of attacks to final population of group. Inoculated 110 ÷ 1,000 = 11.0 per cent. Uninoculated 890 ÷ 4,000 = 22.3 per cent.

True Comparison – Ratio of attacks to person-years of exposure. Inoculated 110 ÷ (300 × 1/4) + (900 × 1/4) + (1,000 × 1/4) = 20 per cent. Uninoculated 890 ÷ (5,000 × 1/4) + (4,700 × 1/4) + (4,100 × 1/4) + (4,000 × 1/4) = 20 per cent." (Hill, 1939 pp. 128–130).

Using the terminology adopted in this book, the risks (number of cases divided by persons at risk) were 11% in the inoculated and 22.3% in the uninoculated. Apparently, inoculation protected. But the period during which cases were ascertained was shorter for the inoculated than it was for those uninoculated, because the inoculation had been done progressively between April and October of the year of observation. Using person-years at the denominator corrected this imbalance and revealed that the rate was 20 per hundred per year, identical in both groups. The valid conclusion was that inoculation is useless.

The important concept was that a risk was always implicitly associated with a period over which it applied. A risk of 20% has a different meaning if it is expressed

over 6 months, one year or ten years. There is no doubt that this was understood before Hill. But Hill's example shows how critical this characteristic of risk can be, especially for group comparisons.

2.4.3. *Incidence density and cumulative incidence*

Olli S. Miettinen, from the Department of Epidemiology and Biostatistics at Harvard School of Public Health, revisited the relation of risk to rate 138 years after Farr in another seminal paper in the history of epidemiologic methods and concepts entitled "*Estimability and estimation in case-referent studies*" (Miettinen, 1976a). The paper addressed a problem very different from Farr's preoccupation with respect to prognosis: it had to do with the relation of case-control (which Miettinen termed case-referent) and cohort studies (see section 3.11).

Miettinen renamed the incidence rate "incidence density", and interestingly, listed as synonyms two of Farr's expressions, "force of morbidity" and "force of mortality". Miettinen also popularized the term "cumulative incidence" instead of "risk". The properties of risks and rates remained those described by Farr, but Miettinen showed that the risk could be expressed as a function of the incidence density (ID). In its simpler formulation:

$$\text{Cumulative incidence}_{(\text{up to time } j)} = \Sigma_{\text{from time } i = 1 \text{ to } j}\, ID_i$$

For example, suppose that the incidence rate of a relatively rare disease (e.g., breast cancer) changes at each year of age and that there is no cohort effect (see section 3.4.3). The risk of a woman to develop breast cancer before age 75 is the sum of the 74 age-specific incidence rates between birth and age 74. In Western societies, this cumulative incidence is about 7%. The formula found in Miettinen's paper (Miettinen, 1976a) allows for the possibility that incidence rates are stable over specific time periods, Δt (e.g., $\Delta t = 5$ for a 5-year risk). In this situation:

$$\text{Cumulative incidence}_{(\text{up to time } j)} = \Sigma_{\text{from time } i = 1 \text{ to } j}\, ID_i \times \Delta t_i$$

Miettinen's innovative concepts have reached a much larger audience than the papers in which he developed them. The original papers can be arduous for someone who is not already familiar with epidemiologic concepts and methods and does not have some mathematical background. Therefore, his concepts have usually been disseminated through the work of people who wrote didactic translations of his ideas. We owe to a group of epidemiologists and statisticians at the School of Public Health of the University of North Carolina and Yale University, Hal Morgenstern, David G. Kleinbaum and Lawrence L. Kupper a paper that translates Miettinen's 1976 "*Estimability*" paper into a more universally accessible prose (Morgenstern et al., 1980).

The paper reminded first that:

> *"(...) the concept of risk requires a specific period referent, – e.g., the 5-year risk of developing lung cancer."* (Morgenstern et al., 1980, p. 97).

When computing the risk, that is, the proportion of all the subjects at the onset who developed the disease during a given period, we assume that all subjects have been followed during the full period. What happens when this condition of complete follow-up is not met? William Farr and Bradford Hill had shown that we could avoid a bias by computing incidence rates based on person-times, instead of risks. Miettinen proposed the following solution: divide the duration of follow-up, t, into short time intervals; compute a risk for each short interval and call it incidence density (ID); sum the incidence densities over all time intervals and you get the cumulative incidence (CI) over the period t. The cumulative incidence is a measure of the risk over period t. Using Miettinen's formula given above, we can compute the cumulative incidence (= risk) as the sum of incidence densities. This measure of risk is not affected by the fact that some observations had incomplete follow-up.

Morgenstern, Kleinbaum and Kupper illustrated the relation of risk (CI) and rate (ID) by the example described in Table 6.

The question is: what is the risk of a 35-year old woman to develop breast cancer before age 55? If we take the 60,000 women in age group 35–39 followed 3 years,

Table 6 – Illustration of the estimation of risk in a dynamic population of 250,000 women free of breast cancer, aged 35 to 55y, followed up for 3 years (on average). Source: Table 1, in (Morgenstern et al., 1980).

Age (yr)	Women at risk [N]	No of incident cases [I]	Person-years [PY = N × 3]	Incidence density[1] (/100,000/yr)	5-year Risk[2] (/100)
35–39	60,000	90	180,000	50	0.250
40–44	70,000	168	210,000	80	0.399
45–49	65,000	215	195,000	110	0.550
50–54	55,000	227	165,000	138	0.686
					20-year Risk[3]
35–54	250,000	700	750,000	–	1.871

[1] Incidence density = I ÷ Person-years.

[2] Estimate of the Δt = 5-year risk for a woman at the beginning of each age category, $R_{\Delta t} \cong 1 - \exp[-ID \times \Delta t]$.

[3] Estimate of the 20-year risk for a 35 year-old woman, $R_{\Delta t} \cong 1 - \exp[-\Sigma_j ID_j \times \Delta t_j] \cong 1 - \Pi_j (1 - R_{\Delta tj})$.

they represent altogether 180,000 person-years (column 4). The incidence density in this age category is therefore [90 ÷ 180,000 =] 50 per 100,000 per year. Now, the risk of developing breast cancer for a women aged 35 before she reaches 40, that is, over a period of 5 years, is obtained, grossly, by multiplying the incidence density by 5 years, that is, 250/100,000 or 0.25% over 5 years (last column). These 5-year risks increase with age. Thus, the 20-year risk for that same woman aged 35 corresponds, grossly, to the sum of the 5-year risks across the four age categories: [0.0025 + 0.00399 + 0.0055 + 0.00686 =] 1.885%, which is close to the 1.871 per 100 obtained using the appropriate formula mentioned in the Table 6. The answer to the question is: the 20-year risk is about 1.9%.

Note that the formula used to compute the cumulative incidences is more complicated than the simple sum of incidence densities, and should be preferred if the disease is not rare. This example underlines the conceptual evolution between Farr and Miettinen, but does not fully reflect the richness of the theory developed underneath.

2.5. Prevalence and incidence

We have seen that *prevalence* measures the accumulation in the population of events (exposures or diseases) that occurred in the distant or recent past, while *incidence* is a predictive statement about cases-to-be in a population still free of the disease. The two concepts are closely related and their relationships have been explored at least under two different perspectives: a) the relation of incidence to prevalence of disease; b) the relation of (excess) incidence to prevalence of exposure.

2.5.1. Disease prevalence divided by incidence

It has been suggested that Farr had made the first description of the relation between prevalence and incidence, as follows:

> *"... in estimating the prevalence of diseases, two things must be distinctly considered; the relative frequency of their attacks, and the relative proportion of sick-time they produce. The first may be determined at once, by a comparison of the number of attacks with the numbers living; the second by enumerating several times the living and the actually sick of each disease, and thence deducing the mean proportion suffering constantly. Time is here taken into account: and the sick-time, if the attacks of two diseases be equal, will vary as their duration varies, and whatever the number of attacks may be, multiplying them by the mean duration of each disease will give the sick- time."* (Cited by Lilienfeld, 1978, p. 515).