

The Relation Between Risk and Incidence Rate

Because the interpretation of risk is so much more straightforward than the interpretation of incidence rate, it is often convenient to convert incidence rate measures into risk measures. Fortunately, this conversion usually is not difficult. The simplest formula to convert an incidence rate to a risk is as follows:

$$\text{Risk} \approx \text{Incidence rate} \times \text{Time} \quad [4-1]$$

For Equation 4-1 and other such formulas, it is a good habit to confirm that the dimensionality on both sides of the equation is equivalent. In this case, risk is a proportion, and therefore has no dimensions. Although risk applies for a specific period of time, the time period is a descriptor for the risk but not part of the measure itself. Risk has no units of time or any other quantity built in; it is interpreted as a probability. The right side of Equation 4-1 is the product of two quantities, one of which is measured in units of the reciprocal of time and the other of which is time itself. Because this product has no dimensionality, the equation holds as far as dimensionality is concerned.

In addition to checking the dimensionality, it is useful to check the range of the measures in an equation such as Equation 4-1. The risk is a pure number in the range $[0,1]$; values outside this range are not permitted. In contrast, incidence rate has a range of $[0,\infty]$, and time also has a range of $[0,\infty]$. The product of incidence rate and time does not have a range that is the same as risk, because the product can exceed 1. This analysis shows that Equation 4-1 is not applicable throughout the entire range of values for incidence rate and time. In general terms, Equation 4-1 is an approximation that works well as long as the risk calculated on the left is less than about 20%. Above that value, the approximation deteriorates.

For example, suppose that a population of 10,000 people experiences an incidence rate of lung cancer of 8 cases per 10,000 person-years. If we followed the population for 1 year, Equation 4-1 suggests that the risk of lung cancer is 8 in 10,000 for the 1-year period (ie, $8/10,000 \text{ person-years} \times 1 \text{ year}$), or 0.0008. If the same rate applied for only 0.5 year, the risk would be one half of 0.0008, or 0.0004. Equation 4-1 calculates risk as directly proportional to both the incidence rate and the time period, so as the time period is extended, the risk becomes proportionately greater.

Now suppose that we have a population of 1000 people who experience a mortality rate of 11 deaths per 1000 person-years for a 20-year period. Equation 4-1 predicts that the risk of death over 20 years will be $11/1000 \text{ yr}^{-1} \times 20 \text{ yr} = 0.22$, or 22%. In other words, Equation 4-1 predicts that among the 1000 people at the start of the follow-up period, there will be 220 deaths during the 20 years. The 220 deaths are the sum of 11 deaths that occur among 1000 people every year for 20 years. This calculation neglects the fact that the size of the population at risk shrinks gradually as deaths occur. If the shrinkage is taken into account, fewer than 220 deaths will have occurred at the end of 20 years.

Table 4-2 describes the number of deaths expected to occur during each year of the 20 years of follow-up if the mortality rate of $11/1000 \text{ yr}^{-1}$ is applied to a population of 1000 people for 20 years. The table shows that at the end of

Table 4-2 NUMBER OF EXPECTED DEATHS OVER 20 YEARS AMONG
1000 PEOPLE WITH A MORTALITY RATE OF 11 DEATHS PER 1000
PERSON-YEARS

Year	Expected Number		Cumulative Deaths
	Alive at Start of Year	Expected Deaths	
1	1000.000	10.940	10.940
2	989.060	10.820	21.760
3	978.240	10.702	32.461
4	967.539	10.585	43.046
5	956.954	10.469	53.515
6	946.485	10.354	63.869
7	936.131	10.241	74.110
8	925.890	10.129	84.239
9	915.761	10.018	94.257
10	905.743	9.909	104.166
11	895.834	9.800	113.966
12	886.034	9.693	123.659
13	876.341	9.587	133.246
14	866.754	9.482	142.728
15	857.272	9.378	152.106
16	847.894	9.276	161.382
17	838.618	9.174	170.556
18	829.444	9.074	179.630
19	820.370	8.975	188.605
20	811.395	8.876	197.481

20 years, about 197 deaths have occurred, rather than 220, because a steadily smaller population is at risk of death each year. The table also shows that the prediction of 11 deaths per year from Equation 4-1 is a good estimate for the early part of the follow-up but the number of deaths expected each year gradually becomes considerably lower than 11. Why is the number of expected deaths not quite 11 even for the first year, in which there are 1000 people being followed at the start of the year? As soon as the first death occurs, the number of people being followed is less than 1000, which influences the number of expected deaths in the first year. As is seen in Table 4-2, the expected deaths decline gradually throughout the period of follow-up.

If we extended the calculations in the table further, the discrepancy between the risk calculated from Equation 4-1 and the actual risk would grow. Figure 4-3 graphs the cumulative total of deaths that would be expected and the number projected from Equation 4-1 over 50 years of follow-up. Initially, the two curves are close, but as the cumulative risk of death rises, they diverge. The bottom curve in the figure is an exponential curve, related to the curve that describes *exponential decay*. If a population experiences a constant rate of death, the proportion remaining alive follows an exponential curve with time. This exponential decay is the same curve that describes radioactive decay. If a population of radioactive atoms converts from one atomic state to another at a constant rate, the proportion of atoms left in the initial state follows the curve of exponential decay. The lower

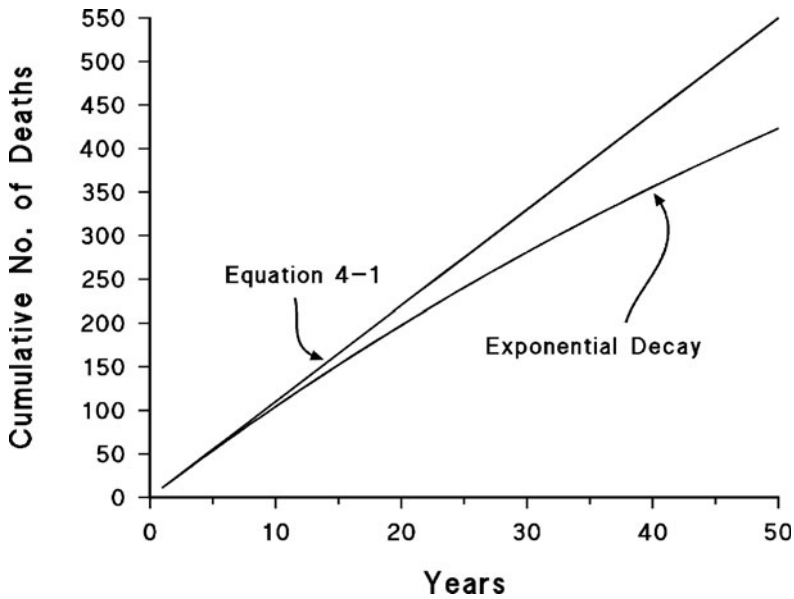


Figure 4-3 Cumulative number of deaths among 1000 people with a mortality rate of 11 deaths per 1000 person-years, presuming no population shrinkage (see Equation 4-1) and taking the population shrinkage into account (ie, exponential decay).

curve in Figure 4-3 is actually the complement of an exponential decay curve. Instead of showing the decreasing number remaining alive (ie, the curve of exponential decay), it shows the increasing number who have died, which is the total number in the population minus the number remaining alive. Given enough time, this curve gradually flattens, and the total number of deaths approaches the total number of people in the population. In contrast, the curve based on Equation 4-1 continues to predict 11 deaths each year regardless of how many people remain alive, and it eventually would predict a cumulative number of deaths that exceeds the original size of the population.

Clearly, Equation 4-1 cannot be used to calculate risks that are large, because it provides a poor approximation in such situations. For many epidemiologic applications, however, the calculated risks are reasonably small, and Equation 4-1 is quite adequate for converting incidence rates to risks.

Equation 4-1 calculates risk for a time period over which a single incidence rate applies. The calculation assumes that the incidence rate, an instantaneous concept, remains constant over the time period. What if the incidence rate changes with time, as is often the case? In that event, risk can still be calculated, but it should be calculated first for separate subintervals of the time period. Each of the time intervals should be short enough so that the incidence rate that applies to it could be considered approximately constant. The shorter the intervals, the better the overall accuracy of the risk calculation, although the intervals should not be so short that there are inadequate data to obtain meaningful incidence rates for each interval.

The method of calculating risks over a time period with changing incidence rates is known as *survival analysis*. It can also be applied to nonfatal risks, but the

approach originated from data related to deaths. The method is implemented by creating a table similar to Table 4–2, called a *life table*. The purpose of a life table is to calculate the probability of surviving through each successive time interval that constitutes the period of interest. The overall survival probability is equal to the cumulative product of the probabilities of surviving through each successive interval, and the overall risk of death is equal to 1 minus the overall probability of survival.

Table 4–3 is a simplified life table that enables calculation of the risk of dying of a motor vehicle injury in a hypothetical cohort of 100,000 people followed from birth through age 85.² In this example, the time periods correspond to age intervals. The number initially at risk has been arbitrarily set to 100,000 people. The life-table calculation is strictly hypothetical, because the number at risk at the start of each age group is reduced only by deaths from motor vehicle injury in the previous age group, ignoring all other causes of death. With this assumption that there are no competing risks, the results are interpretable as risks or survival probabilities that would result if the only risk faced by a population was the one under study. The risk of dying of a motor vehicle injury for each of the age intervals is calculated by taking the number of deaths in each age interval (column 3) and dividing it by the number who are at risk during that age interval (column 2). The survival probability in column 5 is equal to 1 minus the risk for that age category. The cumulative survival probability (column 6) is the product of the age-specific survival probabilities up to that age. The bottom number in column 6 is the probability of surviving to age 85 without dying of a motor vehicle injury, assuming that there are no competing risks (ie, assuming that without a motor vehicle injury, the person would survive to age 85).

Subtracting the final cumulative survival probability from 1 gives the total risk, from birth until the 85th birthday, of dying of a motor vehicle injury. This risk is $1 - 0.98378 = 1.6\%$. Because this calculation is based on the assumption that everyone will live to their 85th birthday except those who die of motor vehicle accidents, it overstates the actual proportion of people who will die in a motor vehicle accident before they reach age 85. Another assumption in the calculation is that these mortality rates, which have been gathered from a cross section of the population at a given time, can be applied to a group of people over the course of 85 years of life. If the mortality rates changed with time, the risk estimated from the life table would be inaccurate.

Table 4–3 LIFE TABLE FOR DEATH FROM MOTOR VEHICLE INJURY FROM BIRTH THROUGH AGE 85^a

Age	Number at Risk	Deaths in Interval	Risk of Dying	Survival Probability	Cumulative Survival Probability
0–14	100,000	70	0.00070	0.99930	0.99930
15–24	99,930	358	0.00358	0.99642	0.99572
25–44	99,572	400	0.00402	0.99598	0.99172
45–64	99,172	365	0.00368	0.99632	0.98807
65–84	98,807	429	0.00434	0.99566	0.98378

^aMortality rates are deaths per 100,000 person-years.

Adapted from Iskrant and Joliet, Table 24.²

Table 4–3 shows a hypothetical cohort being followed for 85 years. If this had been an actual cohort, there would have been some people lost to follow-up and some who died of other causes. When follow-up is incomplete for either of these reasons, the usual approach is to use the information that is available for those with incomplete follow-up; their follow-up is described as *censored* at the time that they are lost or die of another cause.

Table 4–4 shows what the same cohort experience would look like under the more realistic situation in which many people have incomplete follow-up. Two new columns have been added with hypothetical data on the number that are censored because they were lost to follow-up or died of other causes (column 4) and the effective number at risk (column 5). The effective number at risk is calculated by taking the number at risk in column 2 and subtracting one half of the number who are censored (column 4). Subtracting one half of those who are censored is based on the assumption that the censoring occurred uniformly throughout each age interval. If there is reason to believe that the censoring tended to occur nonuniformly within the interval, the calculation of the effective number at risk should be adjusted to reflect that belief.

Point-Source and Propagated Epidemics

An *epidemic* is an unusually high occurrence of disease. The definition of *unusually high* depends on the circumstances, and there is no clear demarcation between an epidemic and a smaller fluctuation. The high occurrence may represent an increase in the occurrence of a disease that still occurs in the population in the absence of an epidemic, although less frequently than during the epidemic, or it may represent an *outbreak*, which is a sudden increase in the occurrence of a disease that is usually absent or nearly absent (Fig. 4–4).

If an epidemic stems from a single source of exposure to a causal agent, it is considered a *point-source epidemic*. Examples of point-source epidemics are food poisoning of restaurant patrons who have been served contaminated food and cancer occurrence among survivors of the atomic bomb blasts in Hiroshima

Table 4–4 LIFE TABLE FOR DEATH FROM MOTOR VEHICLE INJURY FROM BIRTH THROUGH AGE 85^a

Age	At Risk	Motor Vehicle Injury Deaths in Interval	Lost to Follow-up or Died of Other Causes	Effective Number at Risk	Risk of Dying	Survival Probability	Cumulative Survival Probability
0–14	100,000	67	9,500	95,250	0.00070	0.99930	0.99930
15–24	90,433	301	12,500	84,183	0.00358	0.99642	0.99572
25–44	77,632	272	20,000	67,632	0.00402	0.99598	0.99172
45–64	57,360	156	30,000	42,360	0.00368	0.99632	0.98807
65–84	27,204	64	25,000	14,704	0.00435	0.99565	0.98377

^aMortality rates are deaths per 100,000 person-years.