This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/3.0), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.
Published by Oxford University Press on behalf of the International Epidemiological Association *International Journal of Epidemiology* 2012;**41**:1472–1479
© The Author 2012; all rights reserved.
doi:10.1093/ije/dys142

EDUCATION CORNER

Incidence rates in dynamic populations

Jan P Vandenbroucke[1]* and Neil Pearce[2,3]

[1]Department of Clinical Epidemiology, Leiden University Medical Center, RC Leiden, The Netherlands, [2]Faculty of Epidemiology and Population Health, Departments of Medical Statistics and Non-communicable Disease Epidemiology, London School of Hygiene and Tropical Medicine, London, UK and [3]Centre for Public Health Research, Massey University, Wellington, New Zealand

*Corresponding author. Department of Clinical Epidemiology, Leiden University Medical Center, PO Box 9600, 2300 RC Leiden, The Netherlands. E-mail: J.P.Vandenbroucke@lumc.nl

Accepted 30 July 2012

The purpose of the present article is to explain the calculation of incidence rates in dynamic populations with the use of simple mathematical and statistical concepts. The first part will consider incidence rates in dynamic populations, and how they can best be taught in basic, intermediate and advanced courses. The second part will briefly explain how and why incidence rates are calculated in cohorts.

Introduction

The calculation of frequencies of disease is the most basic tool for epidemiology. However, the fundamental concept of the nature of the incidence rate and its calculation in dynamic populations is often not well explained, especially in introductory courses. The concept and the method of calculation were already known to William Farr in the 1850s.[1,2] He thoroughly understood the dynamic population concept and used it imaginatively, e.g. to calculate incidences of nosocomial infections in hospitals.[3] Farr clearly distinguished what he called the 'rate of mortality', which we call today the 'incidence rate' (or one of its synonyms, see Box 1), from what he called 'the probability of death', which is today's 'cumulative incidence' (or 'incidence proportion' or other synonyms, Box 1). He explained in detail the concept of the dynamic population as the basis for the calculation of the 'rate of mortality'.

William Farr's knowledge was subsequently largely forgotten in medicine and epidemiology. It survived, with other generally overlooked scholarship, in demography—a discipline that, like epidemiology, sees Farr as one of its pioneers.[5] The distinction between the two ways of measuring disease frequency was rediscovered in epidemiology in the 1970s.[2,6–9] Still, the concept of dynamic populations as the basis for incidence rate calculations remains often inadequately understood. This hampers not only the understanding of one of the most fundamental measures in epidemiology, but in addition, it hampers the proper understanding of case–control studies. Insight into the calculation of incidence rates in dynamic populations is necessary to understand how the majority of case–control studies are done, and how the odds ratios from such studies should be interpreted, as will be explained in our companion paper.[10]

Box 1 Synonyms for incidence rates and cumulative incidences (risks)

Note that the terms used do not usually make it clear whether incidence rates or cumulative incidences are meant. The origins and early history of the use of different terms has been traced by Turner and Hanley.[4]

For incidence rates:

- Force of mortality
- Force of morbidity
- Incidence density
- Hazard and hazard rate (mostly in statistics).

For cumulative incidences:

- Attack rate
- Case fatality rate
- Lethality
- Risk
- Incidence proportion.

For both (often unspecified which)

- Mortality (rate)
- Morbidity (rate)
- Death rate
- Incidence.

Box 2 Cohort vs dynamic populations

By 'cohort', we concur with the description of 'a group of persons which is determined in permanent fashion or a population, which is determined entirely by a single defining event and so becomes permanent'.[11] Examples are clinical cohorts, such as patients followed up from date of surgery, from date of initiation of a particular drug, or birth cohorts consisting of all persons born in a particular year. The membership of the cohort is fixed by a common event, which is taken as time zero of follow-up. This usage has been long established in epidemiology[12] and is consistent with the definition in the fifth edition of the Dictionary of Epidemiology.[13]

By 'dynamic population', we refer to populations in which the members vary over time; the membership is not fixed. This is a general characteristic of most populations that one commonly thinks of, such as the population of a town or country: 'one can be a member at one time, not a member at a later time, a member again and so on'.[11] This dynamic aspect has also been described as tantamount to observing persons who are in a particular 'state', as long as they are in that state, e.g. as long as they are inhabitants of a town or a country.[12] This usage of 'dynamic populations' is consistent with the definition in the fifth edition of the Dictionary of Epidemiology.[13]

A more complicated example of a dynamic population, which highlights its fundamental characteristics, is an epidemiological study of driving, cellular phone use and accidents. This takes the following form: the observation periods of interest are the periods in which people drive, which they do only intermittently. These observation periods can be divided into subperiods in which the driver is phoning and those in which (s)he is not. What the investigator wants to compare is the incidence rate of accidents while driving and phoning vs while driving and not phoning. In theory, this could be investigated in a cohort study, but the easiest design for this study is a case–control study, as explained in the companion article.[10]

It is important to emphasize that not all authors use these words in the same way. The word 'cohort study' in a publication may indicate either a 'cohort' or a 'dynamic population'; for this reason, some texts refer to the former as 'fixed cohorts' to differentiate them from other types of cohort study (See fifth edition of the Dictionary of Epidemiology for 'fixed cohort').

Advanced technical point

A distinction has been made between 'open' and 'closed' populations, as characteristics that might differ according to the time axis along which the investigator looks at the group (s)he is studying.[11] This distinction is not essential for our basic description of the calculation of incidence rates in dynamic populations and cohorts, but it may become important if the same data are analysed along different time axes. For example, an occupational cohort study can be regarded as 'fixed' (and closed) in that all participants join the cohort on the day they start work and never leave (apart from loss to follow-up, mortality); on the other hand, the study population may be regarded as dynamic (and open) in terms of calendar time, with study participants 'joining' at different times. The study might be analysed for differences in incidence of disease between people who enter at different calendar times, if exposures are judged to vary over time. Moreover, participants may never 'leave' the cohort if they are being followed up indefinitely for cancer incidence, but they may 'leave' if the focus is on workplace injuries, in which case follow-up stops when a participant leaves work.

The purpose of the present article is to explain the calculation of incidence rates in dynamic populations with the use of simple mathematical and statistical concepts. The first part will consider incidence rates in dynamic populations, and how they can best be taught in basic, intermediate and advanced courses. The second part will briefly explain how and why incidence rates are calculated in cohorts.

Dynamic and steady state populations

Basic teaching

Cohorts vs dynamic populations in medicine, demography and epidemiology

When epidemiologists embark on a follow-up study in a group of people, i.e. in a population, that population can present itself to them (or be defined by them) in the following two ways: cohorts or dynamic populations.

Cohorts. In clinical research, most groups of people that are followed up over time present themselves as cohorts. Think of a group of people whom we follow up from surgery until death, e.g. 'the 5-year death rate of a cohort of patients who had surgery for colon cancer in a particular hospital during the year 2005'. The hallmark of a cohort is that its membership is fixed, usually by one defining event (see Box 2 for further details): all those who had surgery during the year(s) in which we accrued patients in a study, belong to the cohort. All are followed up until a particular disease end point or until the closing date of the study. For example if 125 people were operated on in 2005, and 34 die within 5 years after surgery; the 5-year risk of death after colon surgery for cancer

is 34 per 125 or 27 per 100 individuals. In today's epidemiological publications, the commonly used word 'risk' is mostly replaced by more technical terms, such as 'cumulative incidence', or 'incidence proportion' (see below).

The example above assumes that all patients were followed up for a minimum of 5 years after surgery or until death. When using such examples in teaching, depending on the aim of the course, the example can be made more complicated to take into account people who disappeared from the cohort in other ways, such as loss to follow-up or censoring at the end date of the study, which usually leads to the use of either life tables or calculation of person-years of follow-up in cohorts (see below).

Dynamic populations. When demographers think about a population, they think about entities like 'the population of a country during a particular year'. In a particular country in a particular year, say 2005, a number of people are living on the first day of the year, a number of people are living on each subsequent day of the year and a number of people are alive on the last day of the year. These are not all the same people. During the year, some people leave the country or die, other people come to live in the country and babies are born. Such a population is called a 'dynamic population'. The hallmark of a dynamic population is that its members vary, and they are defined by a particular 'state', such as living in a particular country. A person can live in a country for a number of years, and while living in that country, (s)he is a member of that dynamic population; before and after (s)he is not (see Box 2 for further details).

A dynamic population can be understood intuitively as a regiment of a given size in a modern army. Imagine a regiment with a size of 5000 persons. Each time a soldier leaves the regiment, for whatever reason (death, disease, pensioning and so forth), he or she is replaced by a new recruit. The size of the regiment varies slightly from day to day: on some days there are slightly <5000, because the new replacement recruits have not yet arrived; on other days slightly more because the new recruits have arrived before the last day of duty of previous recruits. Even on the battlefield, in today's armies, numbers are sometimes kept constant by flying in new soldiers to replace the dead and wounded. As long as they are members of the regiment, soldiers belong to this dynamic population. Calculations of death rates based on a regiment are straightforward: on average, each day of the year there are 5000 soldiers. Thus, for a year, there are 5000 soldier-years of follow-up. If 63 soldiers die during the year (e.g. in a continuing entrenched war), this would lead to an incidence rate of 63/5000 soldier-years, or 1.3 per 100 soldier-years. This is an incidence rate of death.

In demography, these concepts were already used in the 19th century to calculate population incidence rates. Today, they are still used to calculate death rates in populations of countries, counties or towns; they are also used to calculate 'cancer rates', 'coronary heart disease rates', 'birth rates' or 'marriage rates'. The numerator of such rates is the number of people who developed the condition (e.g. died, developed cancer or gave birth) during a particular year in a country, in a county or in a town. The denominator is *not* the number of people, because people move in and out of the town, county or country, are born and die. The denominator is the 'average' number of people constantly present (alive), multiplied by the amount of time that they are present in the 'risk period' (the particular year, in this example); it is expressed as the number of person-years in the population during that particular year. For example, for a cancer registry of a country in which an average of 2 347 465 women of reproductive age (15–45 years) lived each day of a particular calendar year, and 498 cases of breast cancer occurred in that year, the incidence rate of breast cancer is 498 cases per 2 347 465 women-years or 212 per 1 000 000 women-years (in demographic tables of registries or in vital statistics tables the word 'person-years' is not used, but the concept is referred to as '1 000 000 persons constantly alive'—as if each day of the year there were 1 000 000 persons). Thus, a 'mortality rate', a 'cancer incidence rate', a 'marriage rate' or a 'birth rate' are all incidence rates—they are not cumulative incidences or 'risks'.

Many synonyms exist for the terms that denote risks and rates (see Box 1), which are rooted in the history of these concepts.[4] Because many names are used for the same concepts, it is often not clear from the terminology which is which, and the reader of the literature has to know how the calculations were actually done to understand whether a particular term denotes a rate or a risk. In this article, we will use the term 'cumulative incidence' to denote 'risk' as it is the term most widely used at present, although we should note that the term 'incidence proportion' has been advocated because 'cumulative incidence' has also been used with a slightly different meaning.[11]

A dynamic population can often be seen as in steady state. As a first approximation, for a short period, say a year, dynamic populations of whole countries, counties or towns can be thought about as in 'steady state': on each day, the number of people is more or less the same, although it will fluctuate from day to day. Similarly, the proportion of men and women will be approximately the same on each day, and the age distribution will be roughly the same for all days of the year. Blood group distributions will remain the same (blood group distributions in populations vary only slowly, over decades or even centuries; in general, the genetic make-up of populations remains constant for short periods) and also the number of smokers or the number of vegetarians can be assumed

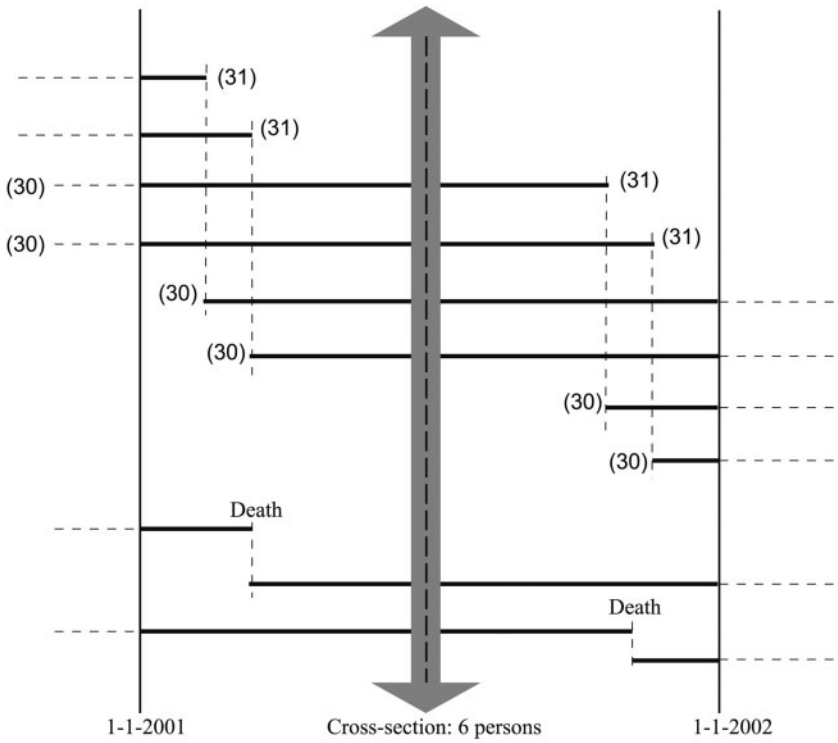

Figure 1 Small scale example, a dynamic population of 30-year-olds that is in steady state during the year 2001. The trajectory of individuals over time is indicated by bold lines. There are two types of 30-year-olds in the course of the year: those who were already 30 years old on 1 January and those who will become 30 years old during the year. The steady state assumes that each time a 30-year-old becomes a 31-year-old, (s)he is replaced by a 29-year-old who becomes a 30-year-old (four new persons in the figure). It also assumes that when a 30-year-old dies (the two bottom lines in the figure), they are replaced by a 29-year-old who becomes a 30-year-old on that day. Thus, on each day there are six individuals who are aged 30 years. On subsequent days, the individuals are different (in total 12 persons contribute to 6 person-years of 30-year-olds). The cross-section in the middle of the year represents the 'average' number of individuals alive during the year. We can calculate the number of person-years in two ways: either by adding all person–time for 30-year-olds or, much simpler, by assessing how many 30-year-olds are present on any day and multiplying this by the time window of 1 year. The incidence rate of death is calculated as two deaths divided by 6 persons-years, conventionally expressed as 33 per 100 person-years (figure adapted from Vandenbroucke *et al.*)[15]

to be in steady state (people stop and start being smokers or vegetarians, and smokers and vegetarians move in and out of town or country, or die). Hence, these subpopulations (women, vegetarians, blood group O carriers, smokers, vegetarians and so forth) can be seen as dynamic subpopulations, that are approximately in steady state for a relatively short period.

Advanced teaching

The underlying concepts about dynamic populations were established in demography in the 19th and first half of 20th century and can be found in classic textbooks of demography, usually in mathematical terms, using calculus (i.e. integration and differentiation).[14] Some epidemiological textbooks cover the principles in depth, but usually in mathematical notation.[8,9,11] The following paragraphs give an account of the underlying principles with the use of only elementary mathematics.

The steady state assumption in more detail

A small-scale example, with only six people, presented and explained in Figure 1 helps to imagine what a dynamic and steady state population of 30-year-olds would look like.

The steady state population assumption uses the idea that people who 'leave', either because they die or because they move out, are constantly replaced by the same type of people. From a demographic point of view, this is less far-fetched than it may seem, at least for short periods. Think about some suburb with which you are familiar: when people move or die, other people come to live in their houses; e.g. when a family with three children moves out of a house, it will be replaced on average by family that is similar, not only in terms of the number of children, but also with regards to socio-economic factors.

The crucial element of this way of thinking is that the population of a suburb in a particular year, say, the year 2005, is *not* the people who lived in that suburb on the 1 January 2005 (which would be the way a clinician would think about a cohort, such as

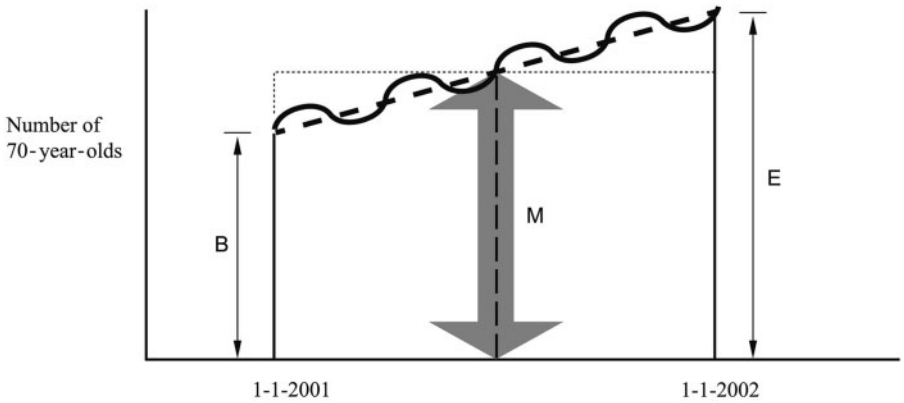

Figure 2 A dynamic population that is *not* in steady state: example of an ageing population. The bold undulating line shows the evolution of the number of 70-year-olds during the year 2001; their number varies from day to day, but grows over time during 1 year. Demographers calculate the 'average number of 70-year–olds' by adding the number of 70-year-olds at the beginning (B) and the end (E) of the year and dividing this sum by two. That is the same as the expected number of 70-year-olds in the middle of the year: $M = (B + E)/2$ under a linear assumption. For a short time, everything can be assumed as approximately linear, even if the number of 70-year-olds will fluctuate; the linear assumption is indicated by the bold dashed line through the undulating line. When you multiply the 'average number of 70-year–olds', i.e. the number in the middle of the year (M) with the time (the year), you get a good approximation of the number of person-years for 70-year-old people during that year, because the surface area of the trapezium (with sides B and E) that you wanted to calculate is the same as the surface area of the rectangle that is completed by the dotted line. This idea goes back to actuarial and demographic theory from the beginning of the 19th century, and it is grounded in elementary calculus: it presents the numerical approximation to integration, i.e. the calculation of an area under a curve (figure adapted from Vandenbroucke *et al.*)[15]

patients after surgery), but the *flow* of people who lived in that suburb throughout the year. That flow is calculated as the number present on average, multiplied by the follow-up time, which yields the 'person-years' (see Figure 1).

What if a population is dynamic but not in steady state?

In real life, dynamic populations are never totally in steady state: towns grow, populations age, neighbourhoods may lose inhabitants or may change with respect to the type of people who live there. However, demographers use a time-honoured and easy solution, which makes the steady state assumption work, even if the underlying dynamic population is not in steady state. We have already used this implicitly in Figure 1 for the simplified example. It consists of taking the estimated population in the middle of the year as an approximation of the 'average' number present for the year. If multiplied by the time of observation (the 'risk period'—1 year in this case), this yields an approximation of the total number of person-years. Figure 2 presents a graph and an explanation of how this looks like for a population that is not in steady state, i.e. an ageing population.

A real life example for the calculations in Figure 2 is as follows. Consider the population of 'males aged 60–64 years in 2001 in The Netherlands', which is a 5-year age group (from the 60th birthday of a person, until the day of his 65th birthday). This population consists of:

(i) All men who were already aged 60–64 years on the 1 January 2001—the 64-year-olds will count up to the date of their 65th birthday, which will be in the year 2001; all others will remain 60–64 during the year and will count for the entire year;
(ii) Plus all 59-year-old men who turned 60 somewhere in between 1 January and 31 December 2001 and stayed 60 years of age for the rest of 2001; these will count from their 60th birthday.

To calculate the number of person-years, we do not need the amount of '60–64-year-old-time' lived by each individual. Instead, from the Central Bureau of Statistics of The Netherlands, we use the following data: on 1 January 2001 there were 368 632 men aged 60–64 years, and on 1 January 2002 there were 375 803 men aged 60–64 years. Thus, on average, there were 372 217.5 men alive each day during 2001. The mortality rate is then calculated as the number of 60–64–year-old males who died in 2001, which is 4648 divided by the amount of person-years of 60–64-year-old males alive in 2001. As the average number was 372 217.5, the number of person-years becomes $372\,217.5 \times 1$.

Using these person–time estimates, the mortality rate will be 4648 per 372 271.5 person-years, which in vital statistics is usually given as a mortality rate

of 12 487 per 1 000 000 person-years (mostly called per 'million constantly alive' in demographic or vital statistics tables). It should be noted that in this example, the numbers alive on 1 January of subsequent calendar years, as reported in vital statistics publications, are themselves interpolations.

Calculating person–time in this manner is tantamount to calculating an 'area under the curve' by numerical integration methods (for more details, treatises on statistics or demography should be consulted). The calculation assumes that for short periods, the increase or decrease of a population can be assumed to be linear (see Figure 2), unless something dramatic happens.

General properties of incidence rates

The cumulative incidence, or risk (also referred to as the incidence proportion),[11,16] calculated from a cohort, is a dimensionless number: people with a particular event are divided by the number of people present at time zero (the common starting point of follow-up, e.g. the day of having surgery). The numerator is contained in the denominator, and the resulting quantity is by necessity always <1. In contrast, the incidence rate can be >1, depending on the units that are used for person–time or when more than one event is counted for an individual. This can be seen if half of the regiment in the above example is killed in one battle in a single day; then the mortality rate on that day is 2500 soldiers/5000 soldier-days. As a day is 1/365.25 of a year (the 0.25 is to correct for leap years), the 'annualized incidence' of death because of this 1-day battle, i.e. if the numbers that would be killed if there would be such a battle each day of the year, is $(2500/5000) \times 365.25$ or 183 per each person-year. The other way in which incidence rates can be >1 is when more than one outcome event is counted, e.g. when the outcome event is a short disease state; e.g. when surveying the incidence of diarrhoea in infants in developing countries, the number of diarrhoeal episodes may easily become >1 per child-year. Counting more than one event in a person is not possible with cumulative incidences, and in some circumstances, it is a distinct advantage of incidence rates.

The reporting of incidence rates that were >1 was the cause of acrimonious accusations of possible fraud against William Farr and Florence Nightingale, who in the 1860s calculated and compared incidence rates of death in hospitals. These rates were sometimes >1—which is, of course, logical, as in those times, more than one person might have died on each hospital bed in a single year. Interestingly, these accusations were rehashed >130 years later, and they needed renewed explanations of the underlying principles.[17,18] In today's times, a hospice or palliative care unit wherein the few beds are in high demand, may also present with an incidence rate of death that is >1.

Although the principles are clear, the incidence rate has occasionally come 'under attack' during the past decades, in particular, because the person-years concept is not understood or because the fact that more than one episode can be counted is not understood.[18,19]

An important caveat with the use of incidence rates is that they are assumed to be constant for the time window in which they are measured. In practice, 10 persons followed up for 100 years will usually show a different incidence rate of death in comparison with 1000 persons followed up for 1 year, although both yield '1000 person-years'. Thus, one should always clearly define the time windows (risk periods) when estimating incidence rates and reflect on whether the proposed time windows, say, for a particular age category of persons for a number of calendar years, is likely to have a reasonably constant incidence rate.[11] If not, follow-up time should be divided into finer strata, to separately estimate mortality rates in different age groups or periods.

On the other hand, this property of incidence rates is at the same time its main advantage: an incidence rate gives insight into the strength of the morbidity or mortality in a dynamic population and is a kind of 'constant' characteristic of that population. This is in contrast to risk calculations from cohorts, which always approach 1 as the follow-up time becomes longer, because 'in the long run, we are all dead'.[20] This is also the reason that incidence rates are sometimes seen as a more basic concept than risks.

Finally, there is an intriguing relationship between incidence rates and life expectancy. In a population that is in perfect steady state, with a constant incidence rate of death, the life expectancy is simply the inverse of the incidence rate. This can be understood because the incidence rate is the number of deaths divided by all years lived, whereas the life expectancy is the number of years lived, divided by the number of persons who lived them.

Person-years calculations in cohorts

Person-years can also be calculated from cohorts. Doll and Hill used person-years as denominators in the 1956 report of their follow-up study of smoking and lung cancer in British doctors.[21] They used an elegant and simple pre-computer-age procedure: they estimated the number of doctors alive in each age category at one particular date of each follow-up year, and then averaged over the successive years, as explained by MacMahon and Pugh.[22] In his influential 1937 textbook on medical statistics, based on a series of educational articles in the *Lancet*, and which was still being reprinted and revised 40 years later, Austin Bradford Hill advocated calculating person–time in cohorts to get rid of the fallacy of 'neglecting the period of exposure to risk'.[23] Unfortunately, he did not introduce the concept as formally as he did with

life tables and survival in cohorts, to which he devoted a full chapter.

Comparisons with the general population

The use of person-years calculations is pivotal to comparing morbidity and mortality in cohorts (with fixed membership) with that in the general population (with variable membership). One application is in occupational health, where incidence rates of diseases in a particular occupational cohort are compared with corresponding incidence rates in the general population. A time-honoured way of making such comparisons is by direct or indirect standardization (indirect standardization is also called the standardized mortality ratio: it applies the incidence rates of disease in the general population to the person-years in the cohort to compare the observed and expected numbers of diseased or deceased persons, either cause-specific or general).[24] Similar calculations were already done in Willam Farr's time,[1] and they are still a standard way to analyse occupational disease and occupational mortality data in today's medical literature, e.g. for radon exposure in uranium mining.[25]

Another example is the comparison of patient cohorts with the general population, for instance, the development of 'secondary malignancies' after a patient is treated for a first malignancy. The frequency of second malignancies is then compared with the baseline rate of the same type of malignancy in the general population. Such comparisons can also be done for patients who have been treated differently, by radiotherapy or chemotherapy in comparison with the general population—e.g. during long-term follow-up of children treated for acute leukaemia.[26]

Although all the above examples are about 'person-years', of course, one can also use person-months, person-days or even person-hours. Person-days were already used by William Farr to calculate the diminishing mortality due to smallpox during the course of the disease.[27] They are still used today, e.g. bed-days are used to calculate the incidence of nosocomial infections in the early days of hospitalization vs later days of hospitalization. Person-days of being at a certain level of anti-coagulation have been used to look for the optimal level of anticoagulation in patients with different indications, i.e. the level with the least thrombosis, but also the least bleeding.[28] Person-hours of being at a certain level of heparinization in an intensive care unit have been used to calculate the optimal level of such anticoagulation therapy during acute haemofiltration.[29]

Relationships between risks and rates

Incidence rates, as calculated based on person-years, can be used to estimate cumulative incidences. For small time windows or when the disease is rare, which is almost always the case when the follow-up time is small, incidence rates and cumulative incidences (risks) that are estimated for the same follow-up period become numerically indistinguishable. This can be seen if one imagines a population of, say, 341 874 adults who are followed up for a single day; if the number of deaths in that day is 23, then the cumulative incidence of death is 23/341 874 or 6.7 per 100 000 persons, whereas the incidence rate of death would be 23/[341 874 − (23/2)] person-days (which amounts to subtracting half of the number of people who died, as an approximation of the number of half-days not lived on that day) and when expressed per 100 000 person-days is also 6.7. If the incidence rate is larger, and/or follow-up time is longer, the calculation involves an exponential assumption, based on principles of calculus, because the same incidence rate will act on an ever-smaller cohort.[7–9,11]

The inverse calculation is also possible, e.g. from randomized trials in which estimates of risk are usually given. An incidence rate can be estimated when the initial number of people in the treatment arms and the average follow-up time in the trial are known (often given in trial reports), as the multiplication of average follow-up time by the number of people in the trial equals the number of person-years of follow-up; the incidence rate is obtained when the number of outcomes in the trial (usually also given) is divided by this number of person-years.

In statistics, the 'hazard' or 'hazard rate' is a peculiar form of incidence rate wherein the follow-up time approaches the limit of zero and becomes infinitesimally small, which is often called an 'instantaneous hazard'. It creates a situation in which there is no more numerical difference between incidence rates and cumulative incidences. It is used, among others, in the proportional hazards model (see our related article on case–control studies).[10]

Importantly, estimation of incidence rates through person-years (or person-days or person-hours) permits, in principle, total flexibility of multivariable analyses, i.e. adding several variables to the analysis by using a Poisson model and slicing up person–time in different ways.

Conclusions

In addition to the concepts of cumulative incidence ('risk') calculation in cohorts, the calculation of incidence rates using person-years in dynamic populations should be taught thoroughly in basic courses of epidemiology. In fact, from a population perspective, incidence rates could be considered the more basic notion than risks. It is important to teach that person–time calculations can be done in dynamic populations and cohorts, whereas risk calculations can only be done directly in cohorts. Moreover, a basic understanding of incidence rates is pivotal to

understanding case–control studies[10] as well as for the understanding of the analyses of many cohort studies and of the basic demographic measures that are used in public health.

Acknowledgements

The authors thank the editors, the anonymous reviewers and Dr J Hanley for their constructive comments that improved the article.

Funding

Jan P Vandenbroucke is an Academy Professor of the Royal Netherlands Academy of Arts and Sciences. The center for Public Health research is supported by a Programme Grant from the Health Research Council of New Zealand.

Conflict of interest: None declared.

References

1 Farr W. *Vital Statistics: A Memorial Volume of Selections of the Reports and Writings of William Farr*. (Reprint of 1885 edition with introduction by Susser M, Adelstein A.) Published under the auspices of the Library of the New York Academy of Medicine. Metuchen NJ: Scarecrow Press, 1975, pp. 392–411, 485–91. [The original 1885 edition is available online at: http://pds.lib.harvard.edu/pds/view/7162583 (12 July 2012, date last accessed)].

2 Vandenbroucke JP. On the rediscovery of a distinction. *Am J Epidemiol* 1985;**121:**627–28.

3 Vandenbroucke JP, Vandenbroucke-Grauls CM. A note on the history of the calculation of hospital statistics. *Am J Epidemiol* 1988;**127:**699–702.

4 Turner EL, Hanley JA. Cultural imagery and statistical models of the force of mortality: Addison, Gompertz and Pearson. *J R Statist Soc A Stat Soc* 2010;**173**(Pt 3):483–99. Additional explanatory slides with additional material available at: http://www.medicine.mcgill.ca/epidemiology/hanley/Reprints/ (12 July 2012, date last accessed).

5 Eyler J. *Victorian Social Medicine: The Ideas and Methods of William Farr*. Baltimore, MD: John Hopkins University Press, p. 1979.

6 Elandt-Johnson RC. Definition of rates: some remarks on their use and misuse. *Am J Epidemiol* 1975;**102:**267–71.

7 Miettinen OS. Estimability and estimation in case-control studies. *Am J Epidemiol* 1976;**103:**226–35.

8 Kleinbaum DG, Kupper LL, Morgenstern H. *Epidemiologic Research*. Belmont: Lifetime Learning Publications, 1982, pp. 96–116.

9 Breslow NE, Day NE. *Statistical Methods in Cancer Research. Vol. 1—The Analysis of Case-Control Studies*. Lyon: IARC, 1980, pp. 42–53.

10 Vandenbroucke JP, Pearce N. Case-control studies: basic concepts. *Int J Epidemiol* 2012;**41:**1480–89.

11 Rothman KJ, Greenland S, Lash TL (eds). *Modern Epidemiology*. 3rd edn. Philadelphia: Lippincott Williams & Wilkins, 2008, pp. 32–50.

12 Miettinen OS. *Theoretical Epidemiology*. New York: Wiley, 1985. pp. 321, 325.

13 Porta M (ed). *A Dictionary of Epidemiology*. 5th edn. Oxford: Oxford University Press, 2008. http://jpkc.fudan.edu.cn/picture/article/189/c4/24/81c086374fd8a31d9-be7208bbb80/eb7e72b0-3b41-4b6b-8b23-168950e0e794.pdf. (12 July 2012, date last accessed).

14 Jordan CW. *Life Contingencies*. Chicago: The Society of Actuaries, 1967.

15 Vandenbroucke JP, Hofman A. *Grondslagen der Epidemiologie*. 6th edn. Maarssen: Elsevier/Bunge, 1999.

16 Pearce N. Classification of epidemiological study designs. *Int J Epidemiol* 2012;**41:**393–97.

17 Iezzoni LI. 100 apples divided by 15 red herrings: a cautionary tale from the mid-19th century on comparing hospital mortality rates. *Ann Intern Med* 1996;**124:**1079–85.

18 Vandenbroucke JP. Continuing controversies over ''risk and rates''—more than a century after William Farr's ''On Prognosis''. *Soz Präventivmed* 2003;**48:**216–18.

19 Windeler J, Lange S. Events per person year—a dubious concept. *BMJ* 1995;**310:**454–56.

20 Keynes JM. *Quotes*. http://en.wikiquote.org/wiki/John_Maynard_Keynes (12 July 2102, date last accessed).

21 Doll R, Hill AB. Lung cancer and other causes of death in relation to smoking. A second report on the mortality of British doctors. *BMJ* 1956;**ii:**1071–81.

22 MacMahon B, Pugh TF. *Epidemiology: Principles and Methods*. Boston, MA: Little, Brown & Co, 1970, p. 228.

23 Hill AB. *Principles of Medical Statistics. Postgraduate Series. Vol. 3*. London: The Lancet, 1937, pp. 128–31.

24 Monson RR. *Occupational Epidemiology*. Boca Raton, FL: CRC Press, 1980.

25 Schubauer-Berigan MK, Daniels RD, Pinkerton LE. Radon exposure and mortality among white and American Indian uranium miners: an update of the Colorado Plateau cohort. *Am J Epidemiol* 2009;**169:**718–30.

26 Pui CH, Cheng C, Leung W *et al*. Extended follow-up of long-term survivors of childhood acute lymphoblastic leukemia. *N Engl J Med* 2003;**349:**640–49. Erratum in: *N Engl J Med* 2003;**349**:1299.

27 Morabia A. *A History of Epidemiologic Methods and Concepts*. Basel: Birkhäuser, 2004, p. 175.

28 Rosendaal FR, Cannegieter SC, van der Meer FJ, Briët E. A method to determine the optimal intensity of oral anticoagulant therapy. *Thromb Haemost* 1993;**69:**236–39.

29 van de Wetering J, Westendorp RG, van der Hoeven JG, Stolk B, Feuth JD, Chang PC. Heparin use in continuous renal replacement procedures: the struggle between filter coagulation and patient hemorrhage. *J Am Soc Nephrol* 1996;**7:**145–50.