Brit. J. prev. soc. Med. (1972), **26**, 132–134

COMPUTING MAN YEARS AT RISK

I. D. HILL

MRC Computer Unit, 242 Pentonville Road, London N1 9LB

A common task in epidemiology (e.g., Doll, 1955; Newhouse, 1969) is the finding of the number of man years experienced by a population in a survey, and multiplying this number by a death rate (usually taken from national figures) to derive an expected number of deaths in the population to compare with the observed number. Usually, however, it is not meaningful to get the man years as just a single number, since death rates for practically all causes of death vary with the age of the individual. It is, therefore, necessary to get the man years in the form of a table, giving each of a number of age groups separately, and to use a different death rate for each cell.

A Simple Example

As a simple example, consider a man born on 16 July 1928, who enters a survey on 5 January 1952. His age at entry will be 23·46 years.

If he leaves the survey on 30 June 1962, his age at exit will be 33·96 years.

If the survey is using five-year age groups running 0–, 5–, 10–, 15–, etc. this man will contribute

1·53 man years to group 20–
5·00 man years to group 25– and
3·96 man years to group 30–

These man years may be added to figures similarly produced for each man in the survey to get a total man years at risk figure for each age group.

For this particular purpose, the reason that each man left the survey is irrelevant. For some the date of exit may be the date of death; others may have emigrated, and in many surveys this is taken as an exit because of the difficulty of keeping track of subjects when they are abroad. For many, however, the date of exit may be an arbitrary date, determined by the research worker, on which the survey is closed, either temporarily or permanently, for analysis.

Even in such a simple case as the one discussed above, it is tedious to do the task by hand, particularly if the number of subjects in the survey is at all large, and an approximation has often been used to simplify the job by counting half a year at entry or exit for each subject rather than working out the correct fraction of a year each time. Given a large number of subjects, this approximation is most unlikely to produce misleading results.

Further Dimensions

Very often, however, the difficulty is increased by the fact that, for many causes of death, it is not sufficient to consider only the ages of the subjects, since there have been secular changes in death rates for each given age.

What is needed, therefore, is not only age groups but calendar groups as well. Working in two dimensions, it is easiest to think diagrammatically, and the Figure shows an appropriate diagram for the same man as was used in the simple example above.

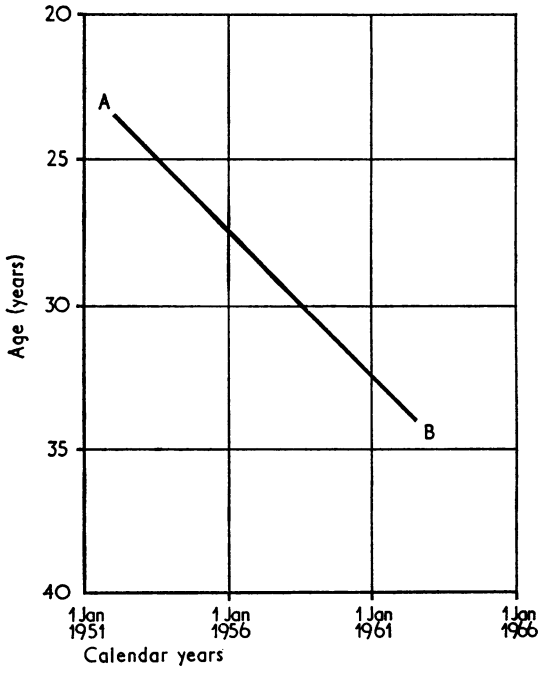

FIGURE. Diagrammatic representation of a man entering a survey at point A, getting older as the calendar years go by, and leaving the survey at point B.

It is assumed here that the calendar scale as well as the age scale is divided into five-year groups,

giving individual cells that are square. This is not a necessary condition; different sized groups could be used for the two scales, leading to rectangular cells, but the particular five-year groups shown here are very convenient in practice since they make the tables of national death rates of Case and Pearson (1957) directly applicable.

The point marked A in the Figure shows the man's age at entry (23·46 years) against the vertical scale and his date of entry (5 January 1952) against the horizontal scale.

Similarly, point B shows his age and date of exit. The line joining these two points must be at 45° to each axis, since he necessarily gets older at the same rate as the calendar years go by.

The number of man years contributed by this man to each cell is proportional to the length of his line in each cell, namely 1·53, 2·46, 2·54, 2·46, and 1·60 man years in each successive cell through which his line passes from point A to point B.

Sometimes even a third dimension is needed; namely, the length of time since entering the survey, sometimes referred to as the 'duration' scale. In this case there are three orthogonal axes, with a line for each subject running at 45° to each of the three. If all three scales have the same width of grouping, the cells now become cubes and the number of man years contributed by each subject to each cell is proportional to the length of his line within the appropriate cube.

Man Years Computer Language

Even in the simple, one-dimensional case the task of forming a man years table by hand is wearisome and liable to error. In two or three dimensions the difficulties are considerably greater, and it is well worth while programming an electronic computer to do the job.

A program called Man Years Computer Language, or Mycl (pronounced to rhyme with cycle) has been written for the purpose. It is called a language since the user specifies his requirements by means of selected English words, together with figures of ages, dates, etc.

This has the advantage not only of making the requirements easy to specify, but also of making it very easy to see precisely what was done if any queries arise at a later date. The need to look back, and to give precise answers to later queries, arises more often than one might guess at first sight. It is a further advantage of the computer approach that such looking back is possible. Those who do the job by hand rarely keep sufficiently detailed records of what was done, and later queries may meet a lack of memory.

Setting out one's requirements in this language is known as writing a Mycl program. This is perhaps slightly to stretch the meaning of the word, since to the computer the Mycl program is merely data, to be operated upon by a program that is already written, but where such data specify not only the numbers to be used but also the operations to be performed, and the order in which they are to be performed, there is precedent for using the word 'program'.

However, although specifying the requirements for Mycl is a form of programming, the language has been written for a special purpose, and the programming job is consequently very much simpler than using a general purpose programming language. Indeed, I hope that it is simple enough for an epidemiologist, without any previous computing experience, to be able to use it.

Versions of Mycl

The original version of Mycl became available in 1966 on the ICL Atlas computer and was written in the Algol computer language. A new version is now available on the XDS 9300 computer, written in machine code.

The Algol version could be relatively easily modified to run on any computer having an Algol complier, if desired.

The two versions are very similar in the facilities provided, although some changes have been made as a result of experience.

It is important to emphasize that the user does not need to know anything of either Atlas Algol or XDS machine code; he only needs to know how to write Mycl instructions.

Main Facilities Available in Mycl

The most important facilities available are:

(1) to specify the dimensions, the scales, and the groupings to be used in the tables, for example:

```
AGE (20 BY 5 (4 GROUPS))
CALENDAR (1/1/1951 BY 5 (3 GROUPS))
```

is a portion of Mycl program that would specify tables to correspond to the diagram of the Figure.

The 'BY 5' specifies groups of a constant five-year width, but it is also possible to specify groups of varying width if desired;

(2) to calculate man years for a particular group of subjects and form a man years table from them. During this operation, checks may be made (at the user's discretion) to look for errors in the data. An example is:

```
DATES (5)
CHECK (AGE AT ENTRY >20)
16/7/28, 5/1/52, 30/6/62*
2/11/22, 2/5/56, 29/2/64;
```

The figure 5, following DATES, specifies which table to put the results into, as the user may have several tables for different groups of subjects. These are indicated by different numbers that are entirely at the user's choice.

Each subject will be checked as read in to make sure that the age at entry is greater than 20, as requested by the user. In addition, certain standard checks, that the figures are possible ones, are made.

Only two subjects are included in the list above, which is unusually short, purely as an example. For each subject three dates are given, of birth, entry, and exit. Each date is given in the British fashion of 'day/month/year', not the American fashion of 'month/day/year'. Only the last two digits of the year figure are given, and 1900 is automatically added; it is possible to specify when 1800 is to be added instead, when relevant.

Finally, each subject's information is terminated by either a star, to indicate that a further subject follows, or a semicolon, to indicate the end of a complete set of subjects;

(3) to print the man years contained in a table, for example:

```
PRINT (5);
```

would print out in a table of rows and columns the man years in each cell of the same table as dealt with above, together with marginal and grand totals. The number 5 indicates that it is the same table;

(4) to read in a set of death rates, for example:

```
RATES (2)
1237, 1120, 1091,
1266, 1062, 992,
1510, 1264, 1195,
2107, 1895, 1840;
```

This would apply to a case where the tables have 12 cells—a rate is given for each. These rates are, in fact, those for males (all causes) in England and Wales for tables with the age and calendar scales specified in the Figure.

Rates per million per annum are assumed unless the user specifies some other denominator. The digit 2 again specifies a table number at the user's choice. If the same number happens to be used for both a man years table and a rates table, this causes no difficulty; Mycl can always distinguish between the two;

(5) to multiply a man years table by a rates table and print out a resultant table of expected deaths, for example:

```
MULTIPLY (MAN YEARS 5 * RATES 2);
```

will multiply, cell by cell, the two tables considered above and print out in a table of rows and columns the expected deaths in each cell, together with marginal, and grand, totals.

SUBSIDIARY FACILITIES Less important, but still useful, facilities are:

(6) to print out titles of the user's choice;

(7) to clear a man years table, i.e., set all its cells to zero, so that it may be used again without the new results being added to man years already there;

(8) to add two or more man years tables together, or to subtract one from another;

(9) to give a name to a table, which is printed out as a heading whenever the printing of the corresponding table is requested;

(10) to switch input channels on the computer, since it can sometimes be useful to have part of a Mycl program on punched cards, for instance, while another part resides on magnetic tape;

(11) to call in additional programming not written in Mycl. This can add to the useful things that can be done, but its use depends upon an understanding of the inner workings of Mycl and is not intended for the beginner.

MYCL MANUAL A Mycl Manual has been written which gives fuller details of all the possibilities. Copies may be obtained on request to the MRC Computer Unit.

SUMMARY

The task of finding the number of man years experienced by a population, for each of a number of age groups and calendar groups, is tiresome and liable to error. A computer program is now available that greatly eases the task.

REFERENCES

CASE, R. A. M., and PEARSON, J. T. (1957). Tables for comparative composite cohort analysis. In: *Cancer Statistics for England and Wales*, 1901–55, pp. 37–99. *General Register Office. Studies on Medical and Population Subjects*, No. 13. H.M.S.O., London.

DOLL, R. (1955). Mortality from lung cancer in asbestos workers. *Brit. J. industr. Med.*, **12**, 81.

NEWHOUSE, M. L. (1969). A study of the mortality of workers in an asbestos factory. *Brit. J. industr. Med.*, **26**, 294.