Five-year age-specific incidence rates

II: The accuracy of calculations of expected numbers of tumours

N. CRAMER

From the School of Economics and Politics, Kingston Polytechnic

I. K. CROMBIE

From the Cancer Epidemiology Research Unit, University of Birmingham

SUMMARY Five-year age-specific incidence rates were shown to produce small but systematic errors in the calculation of the expected number of tumours in a hypothetical but realistic study population. Underestimates occurred at younger ages (under 55) and overestimates at older ages, with a small overestimate (0.22%) overall. Larger errors (up to 12%) were obtained when there was a rapid change in the single-year age structure of the study population. Interpolation between five-year rates will normally produce an inaccurate set of one-year rates. It is shown, with the example of a logarithmic interpolation, that these rates tend to produce errors of similar size to the five-year rates but with a small underestimate overall (0.37%). However, the interpolated rates produced the smaller errors (up to 1%) when the study population age structure undergoes rapid change. A method is suggested for partially correcting the error in the interpolated rates.

Epidemiological studies frequently compare the observed cancer experience of a study group with that of the general population. The expected number of tumours is normally calculated by applying five-year age-specific incidence rates to the accumulated person-years of observation subdivided by age.¹ These studies are expensive and time-consuming and it is clearly worthwhile to consider ways in which the calculation of the expected numbers could be refined. One obvious possibility is the use of one-year rates for five-year rates, but such rates are not available from published data and would have to be derived by some method of interpolation, a lengthy and possibly error-prone procedure.

In this study we investigated the size of the errors arising from the use of five-year rates, comparing the results with those from a defined set of one-year rates. We examined a method of obtaining one-year rates by interpolating between five-year rates, determined the source and size of the errors which this entails, and suggested a possible solution.

Results

THE BASIC DATA

The one-year age-specific incidence rates corresponding to stomach cancer among women in

the West Midlands, described in the companion paper,² were used. The five-year incidence rates which would be observed if the one-year rates had occurred in a defined population (the 1961 England and Wales life table population) were calculated.³ An examination of the literature³⁻⁹ revealed that study populations (or more accurately the person-years at risk) typically exhibit, with advancing age, an initial rapid increase followed by a slow but ever-increasing rate of decrease. Such a study population defined by single years of age is described in Table 1, with the change in size across each five-year age group occurring by four successive equal annual increments.

EXPECTED NUMBERS

The expected numbers obtained using the five-year and one-year rates and the defined study population are shown in Table 1. Three important features emerge: (i) the majority of the tumours occur at ages 50–79, where the progressive increase of the incidence more than outweighs any decline in the size of the study population, thus, errors affecting these ages will tend to predominate in the all-ages figure; (ii) the pattern of the errors produced by the five-year rates consistently underestimates at ages under 55—and overestimates at ages over 60, the direction of the errors corresponding to the change which occurs

Table 1 Comparison of the expected numbers produced by five-year and one-year incidence rates

Age group	Size of mid-one-year group	% Annual change ^a within five-year group	Expected numbers		Error of 5-year rates ^b
			1-year rates E(I ₁)	5-year rates E(I ₅)	
35–	500	40	0.116	0.105	–0.011 (–9.5) ^c
40–	3 000	33.3	1.277	1.183	–0.094 (–7.4)
45–	7 700	11.7	5.473	5.335	–0.138 (–2.5)
50–	9 800	1.0	11.314	11.272	–0.042 (–0.4)
55–	9 550	–1.6	17.411	17.411	0.000
60–	8 500	–2.9	23.617	23.627	+0.010 (0.04)
65–	6 800	–5.9	27.816	27.885	+0.069 (0.2)
70–	4 350	–12.6	25.373	25.576	+0.203 (0.8)
75–	1 900	–23.7	15.330	15.614	+0.284 (1.9)
All ages	—	—	127.727	128.008	+0.281 (0.2)

a. Annual increment as percentage of mid-one-year group.

b. Difference between the expected numbers obtained from the five-year and one-year incidence rates.

c. Error as percentage of one-year rate value.

across the five-year study population groups: an increase led to an underestimate, a decrease to an overestimate. For all ages the errors partially cancel out to produce a small overestimate (0.22%); and (iii) the size of the error, as a percentage, depends on the direction and amount of change across the five-year study population groups, with larger errors occurring when the study population increases. Annual changes in the range +5% to –15% will tend to produce errors of less than 1%.

The use of a five-year rate in the calculation of expected numbers is equivalent to applying five one-year rates (equal in magnitude to the five-year rate) to the one-year population groups. Thus the expected number, E, given by both one-year and five-year rates can be expressed in the form

$$E = \sum_{j=1}^5 N_j \cdot I_j \quad \dots (1)$$

where N_j and I_j are the one-year study population sizes and incidence rates. The I_j for the five-year rate will take a value approximately (see below) that of the central one-year rate (I_3) whereas the I_j of the one-year rates will increase across the age-group. When the N_j increases over the five years (as at ages 35– to 50– in Table 1) extra weight will be given to the larger I_j (that is, I_4 and I_5) of the one-year rates, and the one-year rate expected number will exceed that of the five-year rate. When the N_j decreases, the contribution of the larger I_j of the one-year rates will be reduced and the one-year rate expected number will be the smaller (ages 60– to 75–, Table 1). In the group 55–, the increase in the I_j balances the decrease in the N_j , for this data set, giving identical expected numbers for the one-year and five-year rates.

METHODS OF INTERPOLATION

One-year rates are frequently obtained by logarithmic interpolation between successive

five-year rates (for example,¹⁰). In any such interpolation it is normally assumed that the five-year rate (I_5) is equivalent to the one-year rate of the central year of the five-year period (I_3). But we have recently shown² that the five-year rate differs systematically from this mid-year rate, overestimating the one-year rate at younger ages, and underestimating at older ages. When logarithmically interpolated one-year rates were applied to the study population, they were found to produce small overestimates at the younger ages, and underestimates at the older ages (only some age groups are shown).

Age	40–	50–	55–	60–	70–	All ages
Error from using interpolated one-year rates	+0.017	+0.023	–0.005	–0.041	–0.201	–0.473

Overall, the interpolated rates produce a small underestimate (0.37%), which, for this data set, is larger than the error produced by the five-year rates. In general, five-year rates will produce the smaller errors at ages 50–79 if the study population undergoes a small to moderate decrease across the five-year groups.

ERRATIC STUDY POPULATIONS

Some extreme forms of unusual one-year age structures are described in Table 2, and the ages at which they might occur are shown in Table 3; for example, exponential increase, or decrease, would be near the start, or end, of the study group. The total of person-years in each group was adjusted to correspond with the total numbers in the equivalent age groups analysed in Table 1.

The use of the five-year rates resulted in large errors in the expected numbers, and these were greatest with groups D and B (incidence rates from ages 70– and 75–) (Table 3). The size of the errors corresponded to the amount of change across the group, relative to the size of the central year, rather than to the type of the change. Erratic changes within

Table 2 Examples of five-year groups showing unusual one-year age structure

One-year ages	TYPES OF FIVE-YEAR GROUPS					
	First or last year with zero individuals		Exponential rate of change in size		Erratic changes in size	
	A	B	C	D	E	F
1	—	4 000	485	11 230	5 000	11 000
2	250	2 900	970	5 615	9 500	7 000
3	500	1 800	1 935	2 805	11 000	3 000
4	750	800	3 870	1 400	10 000	4 000
5	1 000	—	7 740	700	7 000	9 000
Total	2 500	9 500	15 000	21 750	42 500	34 000

Table 3 Expected numbers of tumours when the five-year groups have unusual age structure

Nature of one-year age structure	Appropriate age group for incidence rates	Expected tumours using:			Errors from using:	
		True one-year rates	Five-year rates	Interpolated one-year rates	Five-year rates	Interpolated one-year rates
First year zero persons (A) ^a	35—	0.119	0.105	0.120	— 0.014 (— 7.0) ^b	+0.001 (+0.5)
Exponential increase in size (C)	40—	1.351	1.183	1.356	— 0.168 (— 12.4)	+0.005 (+0.4)
Erratic change (E)	60—	23.909	23.627	23.873	— 0.282 (— 1.2)	— 0.036 (— 0.2)
Erratic change (F)	65—	27.673	27.885	27.546	+ 0.212 (+ 0.8)	— 0.127 (— 0.5)
Exponential decrease in size (D)	70—	23.737	25.576	23.576	+ 1.839 (+ 7.7)	— 0.161 (— 0.7)
Last year zero persons (B)	75—	14.721	15.614	14.566	+ 0.893 (+ 6.1)	— 0.155 (— 1.1)
Total		91.510	93.990	91.037	+ 2.480 (+ 2.7)	— 0.473 (— 0.5)

a. These study populations are defined in Table 2.

b. Figures in brackets are the errors as percentages of the true one-year rate.

a group tend to cancel out to produce errors of around 1% (E, F). All these results would be predicted from a consideration of equation (1) and the one-year weighting effects discussed above.

The interpolated one-year rates produced expected numbers which were in general similar to those of the true one-year rates. The largest errors occurred with groups F, D, and B (incidence rates from ages 65—, 70—, and 75—).

Discussion

This study has shown that the errors produced by five-year rates in the calculation of expected numbers can be predicted from the one-year age structure of the study population. The increase in the one-year size invariably seen at the younger ages will result in underestimates, whereas the decrease at the older ages will produce overestimates. These errors will partially cancel out, and overall there will be a small overestimate; because the majority of tumours occur at the older ages, their errors will predominate. Thus, five-year rates will tend to minimise any excess of

observed over expected numbers, and produce conservative estimates of cancer risk.

The size of the error depends on the magnitude of the increase or decrease in the study population, with decreases producing the smaller errors. Progressive single-year changes in the range +5% to –15%, expressed relative to the central year of the five-year group, will produce errors of less than 1%. In contrast, a recent study has shown that the methods of recording and calculating incidence rates could introduce errors of between 1% and 8%.² Since the efficiency of the process of cancer registration may introduce further errors into five-year rates,^{11–13} those errors associated with their use in calculating expected numbers will not normally be a major problem.

The non-equivalence of the five-year rate (I_5) and the central one-year rate of the five-year period (I_3) will introduce errors into one-year rates produced by simple interpolation. The errors in the calculation of expected numbers were of the same order as those produced by five-year rates. Further, because over all ages the interpolated rates tend to give

underestimates (usually at the older ages where most tumours occur) they will increase the apparent size of any observed cancer risk.

The error inherent in interpolation cannot easily be overcome. The relationship between I_0 and I_1 depends on the one-year age structure and incidence rates in the general population, and changes as these change with age.⁸ One solution would be to obtain the one-year population age structure and approximate one-year rates by interpolation, and then calculate 'new' five-year rates from these. The ratio of the observed to the 'new' five-year rates (they will always be found to differ) will give a measure of the error associated with interpolation and could be applied as a correction factor to the interpolated one-year rates. This procedure will only be approximate, particularly because of the difficulty in deriving one-year population structures from the published five-year data.

The conclusion from this work is that five-year rates will give reliable estimates of expected numbers, and that one-year rates obtained by simple interpolation procedures could give larger errors. The single-year age structure of the study population should be examined: only if it contains several age groups across which there are large changes in size will the five-year rates produce large errors and the interpolated rates be more accurate.

I.K.C. was supported by a grant from the Cancer Research Campaign.

Reprints from Dr. I. K. Crombie, A.R.C. Epidemiology Research Unit, Stopford Building (University of Manchester), Oxford Road, Manchester M13 9PT.

References

- ¹Case RAM, Lea AJ. Mustard gas poisoning, chronic bronchitis and lung cancer. *Br J Prev Soc Med* 1955; **9**: 62-72.
- ²Crombie IK, Cramer N. Five-year age specific incidence rates. I Their nature and limitations. *J Epidemiol Community Health* 1980; **34**: 223-8.
- ³Hammond EC. Smoking in relation to death rates of 1 million men and women. *National Cancer Inst Monogr* 1966; **19**: 127-204.
- ⁴Gaffey WR. Epidemiological studies in industry. *JOM* 1973; **15**: 782-5.
- ⁵McMichael AJ, Spirtas R, Kupper LL. Epidemiologic study of mortality within cohort of rubber workers. *JOM* 1974; **16**: 458-64.
- ⁶Robinson TR. Twenty-year mortality of tetraethyl lead workers. *JOM* 1974; **16**: 601-5.
- ⁷Thorpe JJ. Epidemiologic survey of leukaemia in persons potentially exposed to benzene. *JOM* 1974; **16**: 375-82.
- ⁸Goodkin G. Mortality factors in diabetes. *JOM* 1975; **17**: 716-21.
- ⁹Pell S. Mortality of workers exposed to chloroprene. *JOM* 1978; **20**: 21-9.
- ¹⁰Bell R, Case RAM. *Serial Abridged Life Tables: England and Wales 1841-1970. Part II Interpolated Cohort Life Tables*. London: Chester Beatty Research Institute, 1971.
- ¹¹Linden G. The comparability of cancer incidence rates. *Recent Results Cancer Res* 1975; **50**: 141-7.
- ¹²Tuyns AJ. The collection of available information: Possibilities and limitations of international standardisation. *Recent Results Cancer Res* 1975; **50**: 14-19.
- ¹³Muir CS, Waterhouse JAH. Reliability of registration. In: *Cancer Incidence in Five Continents vol III*. Waterhouse J, Muir C, Correa P, Powell J, Davis W, ed. Lyon: IARC, 1976: 45-51.