 Printed in Great Britain
International Journal of Epidemiology 2000;**29**:878–884

Analysing the difference due to risk and demographic factors for incidence or mortality

SA Bashir[a,b] and J Estève[a]

Background From an epidemiological and public health perspective there is an interest in quantifying differences in incidence and mortality between either time points, geographical areas or males and females. We propose a method for splitting such a difference in the number of cases/deaths into three components: (1) those due to risk; (2) those due to population structure (i.e. age distribution); and (3) those due to population size. We also propose graphical methods for presenting the results. Three examples are used to illustrate our methodology.

Keywords Incidence, mortality, risk, demography, rate, cancer

Accepted 24 March 2000

Temporal and geographical trends in disease incidence and mortality provide important information to epidemiologists. However, when it comes to analysing the difference between two given time points, two geographical areas or even between the sexes there is no clear reference as to how this is done in practice with a mathematical foundation.

Methods have been used that do not adjust the population leading to incorrect results.[1,2] Macfarlane *et al.*[2] present the percentage change in the crude rate and the percentage change in the number of deaths for cancers of the upper aerodigestive tract in a variety of different countries. Their aim is to relate these changes to national consumption of alcohol and smoking. However, just looking at the crude rate does not provide a change in risk as structural changes have not been considered. Further they state, 'Numbers of deaths increased in some of the countries with decreasing rates, due to an ageing of the population', but no supporting analysis has been provided. Even if the rate decreases and the population size increases (e.g. doubles or triples) without affecting the structure then we could still see an increase in the absolute number of deaths and this would not be as a result of population ageing. In this situation we would recommend using the method we describe below to calculate the risk and demographic changes. In order to decide whether the structural changes are due to population ageing, one would need to study the population proportions more closely. Engeland *et al.*[1] essentially use the same method as us but they do not adjust the populations to the same size. If the population size and structure does not change the lack of adjustment does not result in large errors. Hence the population size influences the changes due to risk and structure.

In this paper we present a method for partitioning the variation in the number of cases or deaths between two groups (or chronological dates) with respect to demographic variation on the one hand and differences in exposure to risk factors on the other. The demographic component of the variation has to be split itself into that due to variation in population size, which is trivial, and that due to the change in the population structure (i.e. age distribution) which needs more attention.

Methods

Assume that we have the age-group specific number of cases/deaths and population statistics for two 'groups' that are going to be compared, and that the age groups are the same for both groups. Further let group 1 be the *baseline group* (i.e. the reference group) and group 2 the *comparison group* (i.e. this group will be compared to the baseline group). We are interested in partitioning the difference in the total number of cases/deaths between these two groups into three components: (1) differences due to the population size; (2) differences due to the population structure (age distribution); and (3) differences due to the risks.

To eliminate the effect of the population size we start by adjusting the populations so that they comprise of the same number of people (100 000, say) but keep their specific age distribution, i.e. the proportion in each age group are the same as in the total population. This is equivalent to working on the crude rate instead of the number of cases.

Then, we have only to partition the difference in the crude rates between those due to differences in the population structures and those due to the differences in the risks. This is done by comparing the rates in the two groups to an intermediate rate obtained by applying the baseline age-specific incidence/mortality rate to the age distribution of the comparison group.

$$\frac{S_2 - S_1}{S_1} = \frac{S_3 - S_1}{S_1} + \frac{S_2 - S_3}{S_1}$$

where S_1 and S_2 are the crude rates (per 100 000 people) for group 1 and group 2, respectively. S_3 is the 'intermediate' rate obtained when the baseline age-group specific rate is applied to

[a] Service de Biostatistique, Batiment 1M, Centre Hospitalier Lyon Sud, 165 Chemin du Grand Revoyet, 69495 Pierre-Benite, France.

[b] Current address: Amgen Ltd, 240 Cambridge Science Park, Milton Road, Cambridge CB4 0WD, UK.

Table 1 Observed number of deaths due to lung cancer in French males with population numbers and the populations adjusted to 100 000 people

	Deaths		Population		Adjusted population	
Age (years)	1970	1980	1970	1980	1970	1980
0–4	3	1	2 135 500	1 889 000	8615	7180
5–9	0	0	1 955 300	1 889 000	7888	7180
10–14	1	1	2 109 000	2 126 800	8508	8084
15–19	4	1	2 115 800	2 190 100	8535	8324
20–24	4	3	2 228 900	2 142 500	8991	8143
25–29	5	8	1 568 200	2 140 000	6326	8134
30–34	19	35	1 573 800	2 254 500	6349	8569
35–39	75	89	1 680 300	1 580 500	6778	6007
40–44	234	289	1 702 800	1 535 200	6869	5835
45–49	443	711	1 625 700	1 613 900	6558	6134
50–54	514	1331	1 017 700	1 590 400	4105	6045
55–59	1188	1959	1 237 500	1 472 200	4992	5596
60–64	1906	1661	1 235 300	889 000	4983	3379
65–69	2159	2588	1 061 700	996 100	4283	3786
70–74	1811	2843	747 400	883 300	3015	3357
75–79	895	2204	424 400	630 500	1712	2396
80–84	427	1112	240 400	329 100	970	1251
85+	188	422	129 300	158 000	522	601
Total	9876	15 258	24 789 000	26 310 100	100 000	100 000

the comparison group population. S_3 is an age-standardized rate for the baseline group using the population of the comparison group as the standard population (direct standardization[3]). Alternatively, S_3 is also the expected rate of the comparison group considering the baseline group as the standard population (indirect standardization[3]).

The first component on the right-hand side of equation (1) represents the proportional change in the crude rates due to differences in the population structure (i.e. age distribution) and the second component those due to differences in the risk.

The full algebraic formulation is given in the Appendix.

Examples

Differences between time points

To illustrate the idea more clearly we use data extracted from the WHO mortality database for lung cancer in French males. We aim to analyse the change between, say, 1970 and 1980, in terms of differences in demography and risk.

Table 1 shows the traditional age-specific data (grouped into 5-year intervals) for the number of observed deaths and population. It can be seen that the total number of deaths increases from 9876 to 15 258 between 1970 and 1980 (i.e. an increase of 5382 [54.5%] deaths). However, looking at the absolute number of deaths can be misleading as the population increased but not uniformly. The total population increased by 6.1%; by about 37% in the age group 80–84 but it decreased in the age group 35–49 years.

Table 2 gives the calculations needed for quantifying the differences due to risk and demography where

R_1 is the set of expected age-group specific number of deaths in 1970 for a total population of 100 000. For example, for the 50–54 age group (Table 1) we have 2.074 = (514/1 017 700) × 4105. S_1 is the sum of the age-specific R_1s (i.e. S_1 is the crude rate for 1970).

Table 2 Calculations for S_1, S_2 and S_3 as shown in the Appendix

Age (years)	R_1	R_2	R_3
0–4	0.012	0.004	0.010
5–9	0.000	0.000	0.000
10–14	0.004	0.004	0.004
15–19	0.016	0.004	0.016
20–24	0.016	0.011	0.015
25–29	0.020	0.030	0.026
30–34	0.077	0.133	0.103
35–39	0.303	0.338	0.268
40–44	0.944	1.098	0.802
45–49	1.787	2.702	1.672
50–54	2.074	5.059	3.053
55–59	4.792	7.446	5.372
60–64	7.689	6.313	5.214
65–69	8.710	9.837	7.699
70–74	7.306	10.806	8.135
75–79	3.610	8.377	5.054
80–84	1.723	4.227	2.222
85+	0.758	1.604	0.873
Total	39.840	57.993	40.536
	S_1	S_2	S_3

R_2 is the set of expected age-group specific number of deaths in 1980 for a total population of 100 000. For example, for the

Table 3 Analysis of the difference in lung cancer mortality for French males between 1970 and 1980 in terms of risk and demographic factors

Change in	Crude rate	%	No.	%
Risk	17.5	43.8	4327	43.8
Structure	0.7	1.7	172	1.7
Size			882	9.0
Net change	18.2	45.5	5382	54.5

70–74 age group (Table 1) we have 10.806 = (2843/883 300) × 3357. S_2 is the sum of the age-specific R_2s (i.e. S_2 is the crude rate for 1980).

R_3 is the set of expected age-group specific number of deaths in 1980 if the risk was the same as in 1970 for a total population of 100 000. For example, for the 65–69 age group (Table 1) we have 7.699 = (2159/1 061 700) × 3786. S_3 is the sum of the age-specific R_3s (i.e. S_3 is the age-standardized rate for the baseline group (1970) using the comparison group (1980) population as standard).

We are interested in splitting the difference between the total expected number of deaths in 1980 (S_2) and 1970 (S_1) for two populations of 100 000 (i.e. we want $S_2 - S_1$). We can break the simultaneous change in rate and structure into two steps. Our baseline (or starting point) is the total expected number of deaths in 1970 in a population of 100 000.

Step 1 hold the rate constant and change the population to that of 1980. This gives us the total expected number of deaths in 1980 (for a population of 100 000) if the risk had not changed since 1970 (S_3). The difference between this and the baseline is due to structural changes in the population (i.e. $S_3 - S_1$).

Step 2 change the rate to that of 1980. This gives us the total expected number of deaths in 1980 for a population of 100 000. The difference between this and the first step is due to changes in risk (i.e. $S_2 - S_3$).

Hence this means that the main interest focuses on the change in crude rate between 1970 and 1980 (i.e. $S_2 - S_1$). It is more interesting to look at the proportion (or percentage) increase compared to the baseline of 1970 and to get this we simply divide by S_1 (i.e. $[S_2 - S_1]/S_1$).

Table 3 (using the results from Table 2) gives the breakdown in the change for lung cancer mortality in French males between 1970 and 1980. We can see that the crude rate increases by 18.2 per 100 000 population between 1970 and 1980, i.e. an increase of 45.5% ([18.2/39.0] × 100%). Of this, 17.5 (58.0 – 40.5) per 100 000 population was due to the change in risk, i.e. an increase of 43.8% ([17.5/39.0] × 100%) and 0.7 (40.5 – 39.8) due structural changes in the population, i.e. an increase of 1.7% ([0.7/39.8] × 100%). In terms of the absolute number of deaths, there was an increase of 5382 deaths of which 4327 (9876 × 43.8%) were due to the increased risk, 172 (9876 × 1.7%) were due to structural changes and the remaining 882 due to changes (i.e. an increase) in the size of the population.

This increase of 882 deaths due to the difference in the size of populations may at first appear to be misleading as it represents an increase of 9.0% whereas we stated above that the population had only increased by 6.1%. The extra 2.9% comes from the fact the structural and risk changes also affect the increase in the population hence the 45.5% increase in the crude rate has to be applied to the 'extra' 6.1% men in the population (i.e. 9.0% = 6.1% × 1.455).

Table 4 Observed number of deaths due to lung cancer in 1990 for males in Denmark and the UK

	Deaths		Population	
Age (years)	Denmark	UK	Denmark	UK
0–4	0	0	149 600	1 967 700
5–9	0	0	136 400	1 870 400
10–14	0	0	162 000	1 761 500
15–19	1	0	188 200	2 01 1300
20–24	0	2	203 200	2 321 200
25–29	1	1	205 900	2 388 700
30–34	1	12	190 400	2 054 700
35–39	3	58	188 000	1 884 800
40–44	23	202	207 600	2 062 200
45–49	43	450	184 600	169 000
50–54	84	893	143 400	1 550 800
55–59	175	1770	124 900	1 450 200
60–64	277	3493	117 600	1 394 500
65–69	425	5254	111 100	1 310 100
70–74	472	5083	89 600	930 500
75–79	374	4932	67 800	725 200
80–84	204	3246	39 400	414 300
85+	97	1528	23 100	219 000
Total	2180	26 924	2 532 800	28 013 100

Table 5 Analysis of the difference in lung cancer mortality for males in 1990 between Denmark and the UK in terms of risk and demographic factors

Change in	Crude rate	%	No.
Risk	10.8	12.6	274
Structure	–0.8	–0.9	–20
Size			24 490
Net change	10.0	11.7	24 744

Differences between geographical areas

Table 4 shows the number of observed deaths due to lung cancer for males in 1990 for Denmark and the United Kingdom (UK). Here it can be seen more clearly that looking at the increase in absolute number of deaths is not relevant as the population in the UK is about 11 times the size of that in Denmark.

Table 5 shows that there is a 11.7% difference in the crude rate between Denmark and the UK (note that Denmark is being used as the baseline group). The risk in the UK is 12.6% greater than that in Denmark, i.e an extra 10.8 deaths per 100 000 population in the UK due to risk. However, due to differences in the population structure, there is a –0.9% difference in the crude rate between Denmark and the UK, i.e. 0.8 less deaths per 100 000 population in the UK. In terms of absolute numbers this means that if the UK had the same size population as

Table 6 Observed number of deaths due to lung cancer in 1990 in Ireland for males and females

	Deaths		Population	
Age (years)	Females	Males	Females	Males
0–4	0	0	135 900	142 800
5–9	0	0	160 600	173 900
10–14	0	0	169 700	173 900
15–19	1	0	166 300	173 800
20–24	0	0	125 000	135 600
25–29	0	0	117 000	115 300
30–34	0	1	121 400	121 600
35–39	2	3	114 700	118 500
40–44	4	12	110 300	115 700
45–49	12	18	88 700	90 700
50–54	23	45	77 500	81 200
55–59	39	81	68 800	69 500
60–64	67	131	70 300	65 400
65–69	70	200	69 800	59 800
70–74	102	204	60 200	49 900
75–79	105	205	48 800	35 200
80–84	43	103	29 200	17 300
85+	17	41	19 900	8500
Total	485	1044	1 754 100	1 748 600

Table 7 Analysis of the difference in lung cancer mortality between Irish males and females in 1990 in terms of risk and demographic factors

Change in	Crude rate	%	No.
Risk	36.0	133.5	648
Structure	–4.9	–17.6	–85
Size			4
Net change	32.1	115.9	559

Denmark in 1990 for males then there would have been an extra 274 deaths due to risk and 20 less deaths due to the different population structures. Hence the remaining 24 490 deaths are due to the larger population size in the UK.

Difference between sexes

Table 6 shows the number of observed deaths due to lung cancer for Irish males and females in 1990. Here the populations are of similar size but the males have a much larger number of deaths.

Table 7 shows that there is a 115.9% difference in the crude rate between males and females (note that females are being used as the baseline group). The risk in the males is 133.5% greater than that in females, i.e. an extra 36.9 deaths per 100 000 population in the males due to risk. However, due to differences in the population structure, there is –17.6% difference in the crude rate between males and females. i.e. 4.9 less deaths per 100 000 population in the males. In terms of absolute numbers this amounts to 648 more deaths due to risk, 85 less deaths due the population structure and 4 less deaths in the male population.

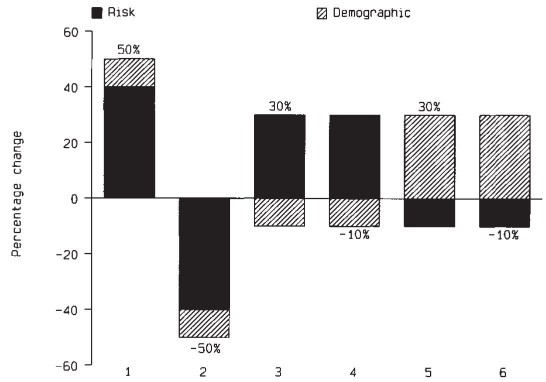

Figure 1 A bar chart for the six possible situations when describing the overall change in the crude rate or the absolute number

Graphical presentation

The graphical presention of these results can get quite complicated and we propose a method which should make it easier to interpret. Firstly, we will only deal with two components, the first is the risk component and the second is for the demographic component. If we are interested in presenting the change in the crude rate then the demographic component will just represent the change in population structure. However, when it is relevant to present the difference in number then the demographic component will represent the change in the population size and structure.

Figure 1 is used to illustrate how we can present the differences under all possible scenarios. There are two situations with a total of six scenarios. In the first situation both the risk and demographic differences could be of the same sign (leading to two possible scenarios as represented by bars 1 and 2 in Figure 1) or they could be of opposite signs (leading to four possible scenarios as presented by bars 3–6 in Figure 1). Below we explain how to interpret the change as presented in Figure 1 (bearing in mind the comments about the demographic component above):

Bar 1 We have a net change of +50% in the crude rate (absolute number) of which +10% is due to demographic factors and +40% due to risk.

Bar 2 We have a net change of –50% in the crude rate (absolute number) of which –10% is due to demographic factors and –40% due to risk.

Bar 3 We have a net change of +30% in the crude rate (absolute number) of which –10% is due to demographic factors and +40% due to risk. Note that the +40% change due to risk is represented by the total length of the bar.

Bar 4 We have a net change of –10% in the crude rate (absolute number) of which +30% is due to risk and –40% is due to demographic factors. Note that the –40% change due to demographic factors is represented by the total length of the bar.

Bar 5 We have a net change of +30% in the crude rate (absolute number) of which –10% is due to risk and +40% due to demographic factors. Note that the +40% change due to demographic factors is represented by the total length of the bar.

Bar 6 We have a net change of –10% in the crude rate (absolute number) of which +30% is due to demographic factors and –40% due to risk. Note that the –40% change due to risk is represented by the total length of the bar.

Further if the bars did not have the net change indicated, it would not be possible to work out whether Bars 3 and 4, and Bars 5 and 6 represented an overall positive or negative difference from the graphs alone. A marker must be added to indicate the end point of the sum of the two components (i.e. the net change). We could use a thick line for this purpose instead of giving the net change in figures.

Bars 3 to 6 can be difficult to interpret initially but if one thinks of starting at zero and then going in the **opposite** direction of the net change this simplifies the interpretation. So, for example, if we look at bar 4 again: we start at zero and firstly there is a 30% increase in risk (i.e. going in the opposite direction to the net change) which is offset by a –40% change in demographic factors (i.e. now starting at +30% and going down by 40%) leading to a net change of –10%.

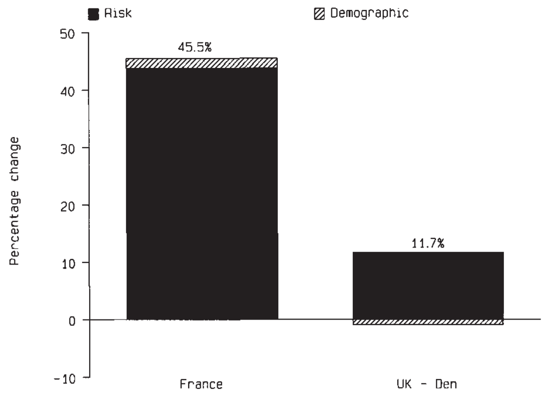

Figure 2 Bar 'France' represents the change in the crude rate for lung cancer in French males between 1970 and 1980. Bar 'UK–Den' represents the difference in the crude rate for lung cancer between Denmark and the UK in 1990

Example

We use the results from 'Differences between time points' and 'Differences between geographical areas' described earlier under 'Examples' to illustrate the use of these graphs (Tables 3 and 5). We will only look at the differences in the crude rate and hence the difference due demographic factors will be due to the population structure only. The results are shown in Figure 2.

The first bar shows the difference in the crude rate for lung cancer in French males between 1970 and 1980 (Table 3). Here there is an overall change of 45.5% of which 43.8% is due to the increase in risk and 1.7% due to the population structure. The second bar shows the difference in the crude rate for lung cancer between Denmark and the UK in 1990 (Table 5). Here there is a net increase of 11.7% of which there is a 12.6% increase in the risk and a 0.9% decrease due to the population structure.

Discussion

We have presented a method of analysing differences in disease incidence and mortality due to risk and demography between two groups. These groups may be two points in time, two geographical areas or both sexes. We used a two-step procedure to breakdown the difference. Starting from a baseline, in the first step we change the (adjusted) population (to a size of 100 000) to that of our comparison group (also adjusted to a size of 100 000) which gives us the structural differences in the population (i.e. differences in the age distribution). In the second step we change the rate to that of the comparison group resulting in the difference due to risk and the remaining number are due to the differences between the size of the two populations.

Care must be taken in presenting the differences in disease incidence or mortality as we can present this difference in terms of the crude rate or absolute number. Although we would advocate the use of the former it is very difficult to get away from the latter. From the public health perspective the absolute number is more useful. However, we must be careful not to misrepresent the true situation if there is a large difference between the two comparison populations. For example, it is quite plausible that the comparison population is many times larger than the baseline population (even between two time points, e.g. developing countries) but at the same time the risk is decreasing. Here we could still see an increase in the absolute number but we should be cautious as more importantly the risk is decreasing.

We also proposed a graphical method for the presentation of the two components (i.e. risk and demographic) using bar charts. However, we would advocate the use of tables ahead of any graphical presentation but at the same time concede there are probably situations in which one needs to present the information graphically.

In this paper we used the raw data in quantifying the difference in lung cancer mortality with respect to demographic and risk factors. However, we would recommend that the data are smoothed before doing such analysis. For example, before comparing two time points we could have done an analysis of time trends using age-period-cohort modelling.[4,5] Here the comparison would have been made using the fitted values. Similarly, one could do some form of spatial smoothing before comparing two geographical points.[6]

Looking at two points may not be useful or it may not paint a clear picture of what is actually happening. We would suggest that multiple comparisons are made to see how the demographic and risk factors change. For example, looking at our example of lung cancer in French males, it would be more informative to look at the evolution of change. Figure 3 and Table 8 show the yearly change between 1970 and 1990 (using 1970 as a baseline). Here we can see that the risk increases faster between 1970 and 1978 compared to between 1979 and 1990. An important feature that would have been missed is that the gap between the net change and the change due to risk is increasing. This means that the changes due to the population structure are increasing (as can be seen from Figure 3 and Table 8). We could have also presented these results as described in 'Graphical presentation'.

One could also look at the change over time when comparing two geographical areas or even both sexes. Finally, a truncated age distribution could be used when making comparisons (e.g. 30–64 years).

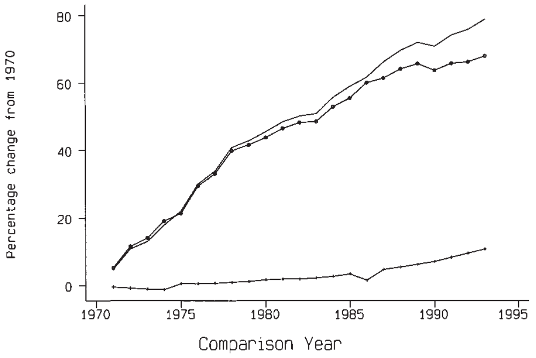

Figure 3 Changes in the crude mortality rate for lung cancer in French males since 1970. • = risk changes; + = structural changes; —— = net change

Table 8 Percentage change in the crude rate for lung cancer in French males compared to the baseline year of 1970

Comparison year	Risk changes	Structural changes	Total change
1971	5.3	–0.4	4.9
1972	11.6	–0.7	10.9
1973	14.1	–1.0	13.1
1974	19.1	–1.2	18.0
1975	21.5	0.6	22.1
1976	29.5	0.6	30.1
1977	33.1	0.7	33.8
1978	39.9	1.0	40.9
1979	41.6	1.2	42.8
1980	43.8	1.7	45.6
1981	46.5	1.9	48.5
1982	48.3	1.9	50.2
1983	48.6	2.3	50.9
1984	53.0	2.7	55.7
1985	55.5	3.5	59.0
1986	60.1	1.6	61.7
1987	61.4	4.8	66.2
1988	64.1	5.5	69.7
1989	65.7	6.3	72.0
1990	63.8	7.1	70.9
1991	65.8	8.4	74.2
1992	66.3	9.6	75.9
1993	68.0	10.9	78.9

Acknowledgements

SAB was funded by the European Union's 'European Network of Cancer Registries' which is part of 'Europe Against Cancer'. This work was completed whilst SAB was at the Service de Biostatistique, Batiment 1M, Centre Hospitalier Lyon Sud, 165 Chemin du Grand Revoyet, 69495 Pierre-Benite, France.

References

[1] Engeland A, Haldorsen T, Tretli S *et al*. Prediction of cancer mortality in the Nordic countries up to the years 2000 and 2010. *Acta Path Microbiol Immunol Scand* 1995;**103(Suppl. 49):**148.

[2] Macfarlane GJ, Macfarlane TV, Lowenfels AB. The influence of alcohol consumption worldwide trends in mortality from upper aerodigestive tract cancers in men. *J Epidemiol Community Health* 1996;**50:**636–39.

[3] Breslow NE, Day NE. *Statistical Methods in Cancer Research, Vol. II: The Design and Analysis of Cohort Studies*. Lyon: IARC Scientific Publications, 1987.

[4] Clayton D, Schifflers E. Models for temporal variation in cancer rates. I: Age-period and age-cohort models. *Statistics in Medicine*, 1987;**6:**449–67.

[5] Clayton D, Schifflers E. Models for temporal variation in cancer rates. II: Age-period-cohort models. *Statistics in Medicine*, 1987;**6:**469–81.

[6] Estève J, Benhamou E, Raymond L. *Statistical Methods in Cancer Research Volume IV: Descriptive Epidemiology*. Lyon: IARC Scientific Publications, 1994.

Appendix

Our aim is to partition the difference in the number of cases (deaths) for two given comparison groups. This difference can be split into differences in risk and those due to differences in demographic factors (i.e. population size and structure).

Assume that we have two comparison groups 1 and 2 for which we have the incidence (or mortality) and populations data by age groups. Further we assume that there are an equal and corresponding number of age groups for both comparison groups, say N.

Let us assume that we have C_1 and C_2 cases/deaths and a total population of P_1 and P_2 in groups 1 and 2, respectively. The relative difference between C_1 and C_2 can be expressed as

$$\frac{C_2 - C_1}{C_1} = \frac{C_2}{C_1} - 1$$

which is the quantity to be analysed. It can be split into the relative difference between the rates.

$$S_2 = \frac{C_2}{P_2} \text{ and } S_1 = \frac{C_1}{P_1}$$

in noting that and the relative difference between P_2 and P_1.

$$\frac{C_2}{C_1} - 1 = \frac{S_2}{S_1}\frac{P_2}{P_1} - 1$$
$$= \left(\frac{S_2}{S_1} - 1\right) + \left(\frac{P_2}{P_1} - 1\right) \times \frac{S_2}{S_1}$$

It is therefore sufficient to analyse $(S_2/S_1) - 1$. Note that the second term above reflects that the change in the population size generates a change in the number of cases which depends on the changes in the crude rate.

This quantity (i.e. $[S_2/S_1] - 1$) is split into a component due to the differences in risk and a component due to differences in the population structure (i.e. age distribution).

Let use define

$$\lambda_{ix} \text{ rate in age group } x \text{ for group } i \quad (2)$$

$$\omega_{ix} \text{ proportion of population in age group } x \text{ for group } i \quad (3)$$

where *i* is 1 or 2. We will be using group 1 as the baseline group for comparison.

We are interested in analysing the difference in the crude rate between groups 1 and 2. Let S_1 and S_2 be the crude rates in comparison group 1 and 2, respectively. So we want to analyse the relative difference in the crude rate, i.e.

$$\frac{S_2 - S_1}{S_1}$$

in terms of risk and population structure (i.e. age distribution). Using (2) and (3), let S_3 be $\Sigma_x \lambda_{1x} \omega_{2x}$ that is the rate in group 1 applied to the population proportion in group 2. We have

$$\frac{S_2 - S_1}{S_1} = \frac{S_2 - S_3}{S_1} + \frac{S_3 - S_1}{S_1}$$

which translates into

$$\frac{\Sigma_x(\omega_{2x}\lambda_{2x} - \omega_{1x}\lambda_{1x})}{\Sigma_x \omega_{1x}\lambda_{1x}} = \frac{\Sigma_x \omega_{2x}(\lambda_{2x} - \lambda_{1x})}{\Sigma_x \omega_{1x}\lambda_{1x}} + \frac{\Sigma_x \lambda_{1x}(\omega_{2x} - \omega_{1x})}{\Sigma_x \omega_{1x}\lambda_{1x}}$$

The first component on the right-hand side represents the proportion of the difference in the crude rate between groups 1 and 2 due to the differences in the population structure and the second component represents the proportion due to differences in the risk.