The Reality of Spatial Variations of Morbidity and Mortality

N. D. McGLASHAN

SUMMARY

Attention is drawn to the inadequacy of using only variations of incidence rates by area as an analytical method in geographical pathology. Problems are illustrated from recent South African literature. Since widely different sizes of population-at-risk produce incidence rates that are not of comparable reliability, advances in probability mapping can appropriately be applied to define significant distributions of medical conditions in South Africa. Both a hypothetical and a real-life illustration of an improved method of analysis are described, and the pressing need for a South African national registry of cancer is emphasised.

S. Afr. Med. J., **48**, 1621 (1974).

More than most countries, South Africa has superb opportunities for studying variations of disease occurrence through space. These opportunities stem from the wide variety of geographical conditions within the Republic's borders, whether they are physical variations of conditions such as temperature, humidity, rainfall and soils, or socio-economic factors such as poverty, literacy, diet and similar locally-varying human customs.

Medical scientists have for many years mapped distributions of cases of disease or death and used the map as a research tool and as a basis upon which to speculate about the causes of disease. Such a method of forming an hypothesis is the critical first step of geographical pathology.[1] The disease map has been in wide use, particularly since 1849 when John Snow demonstrated the water-borne origin of cholera by means of map analysis.[2]

Less widely known, and far less widely used, are modern applications of probability theory to the task of defining meaningful distributions of medical conditions. Yet this statistical advance by medical geographers in the last dozen or so years is so basic and, at the same time, in essence so simple to use, that it deserves greater attention among epidemiologists.

The method is designed to deal with two common difficulties which are often encountered in related circumstances. The first is the question of comparing incidence rates between areas of widely differing population size. A common form of this problem, for example, is to find that urban areas have huge populations-at-risk by comparison with neighbouring rural units of area. One may, rightly, hesitate to make direct comparison of incidence rates between such dissimilar population units.

Secondly, it is obvious that under real world conditions disease cases will not be evenly spread across the landscape *pro rata* to population, viz. at a constant rate of incidence. An element of chance will ensure that some degree of irregularity will occur without any need of attempts to 'explain' the resulting distribution pattern. Some parts of an incidence rate map will show clusters of cases which give a higher incidence, and some will show lower incidence merely by factors which occur at random. One would not attempt to explain why one cricket captain won the toss three times running, but one *would* be entitled to some private suspicion at a succession of thirty lucky calls! Just at what point is explanation to be called for? Ideally the purpose is to distinguish mere chance high (or low) incidence areas from those so extremely high (or low) that they reflect great improbability of occurring thus by chance. If the distribution is not the result of chance, then and only then is explanation appropriate. In the cricket example, perhaps a biased coin . . .

In 1959 Choynowski introduced a paper describing 'Maps based on probabilities'.[3] He considered data on brain tumours in a region of southern Poland and showed, against the Poisson distribution, that most of the spatial variations of incidence rates were due, or could have been due, to chance. A similar testing of which areas within South Africa had more (or fewer) cases than could be expected by chance was presented in 1965 by Oettlé[4] for several sites of cancer with maps designed by McGlashan. A similar technique was again presented by Harington and McGlashan in 1973.[5] Several examples of the use of the probability theory in defining the significance of medical distributions are also included in a recent collection of medical-geography papers.[6]

Since no mathematical expertise is required and the epidemiological gain in terms of the meaning of the map is great, simple examples may assist. Firstly, to illustrate the question of random variations, consider the hypothetical mixed secondary school with an outbreak as recorded in Table I. With a crude outbreak incidence in the school of 4%, can any useful distinction be drawn on the basis of the age or sex of the cases? Incidence (column 5) is apparently markedly higher in the older classes and among the girls. This *might* lead to inquiry about causative factors in those groups of pupils. This, however, would not be justified. If the school rate (4%) is assumed to occur evenly, column 6 shows that each form would expect one single case. The numbers of cases expected (column 6) never fall outside the 95%

University of Tasmania, Hobart, Tasmania
N. D. McGLASHAN, M.A., PH.D.

Date received: 14 March 1974.

TABLE I. HYPOTHETICAL OUTBREAK IN A SCHOOL

(1) Class	(2) Average age	(3) No. in class	(4) Observed cases	(5) Incidence rate %	(6)* Expected cases	(7) Observed: expected	(8)+ $P<0,05$ Poisson limits	(9) Significance
A	17	25	3	12	1	3:1	0,86 - 8,77	Nil
B	16	25	0	0	1	0:1	0 - 3,69	Nil
C	15	25	0	0	1	0:1	0 - 3,69	Nil
D	14	25	1	4	1	1:1	0,03 - 5,57	Nil
Total	15,5	100	4	4	4			
Boys	15,75	70	1	1,43	2,8	1:2,8	0,03 - 5,57	Nil
Girls	15	30	3	10	1,2	3:1,2	0,86 - 8,77	Nil

* Assuming that the school incidence rate (4%) is applied to each class and to each sex.
\+ Poisson tables give a range of limits of observed values for any expected value.

probability level of the Poisson distribution (column 8). Considered by sex (lower part of Table I), the one case among the boys is not outside the 0,03 - 5,57 limits that would be liable to occur *by chance*. The girls have more than twice as many cases observed as would be expected if the school's incidence rate were applied *pro rata* (3 observed: 1,2 expected) but this is still not beyond the limits (0,86 to 8,87) of what could occur by chance. Thus the initial assumption of an even rate in the school is not shown to be disproved. Prophylactic measures or aetiological research should therefore be applied throughout the school, not merely to those sectors with apparently higher incidence. In passing, it may be observed that two widely different populations-at-risk (70 boys against 30 girls) have been validly compared, not by incidence rate, but by level of Poisson significance.

Possibly at risk of appearing invidious, but in order to illustrate these points constructively, two recent South African papers may be mentioned. Both studies were undertaken to identify and then to define on maps areas of higher or lower incidence of oesophageal carcinoma in the Transkei[8] and the Glen Grey district of the Ciskei respectively.[9] Both analyse cancer data at the scale of unit of area of the *location*, the next subdivision of size below the magisterial district. In the Transkei there are 952 locations, and in Glen Grey 22 locations. Location population sizes vary from a few hundred to several thousand, and subdivision by sex further diminishes the size of each unit of population-at-risk. In any location there was a mere handful, usually under a dozen, but often only two or three cases of oesophageal carcinoma in several years' recording. For example, Glen Grey location 22[10] recorded 8 cases (male and female) among a population of 6 274 in 8 years and 3 months; one case per year with an annual incidence rate of 15/100 000. Any error of initial data, whether diagnostic or geographical, is multiplied by 16 times (that is 100 000/6 274) on account of population size alone. That incidence figure of 15/100 000 is then mapped as if it were comparable with 21 other locations with varying-sized multiplications of any errors which may initially have existed (for any reason) in *their* medical records.

Yet the purpose of the exercise is to *define meaningful spatial variation*. In Table II the Glen Grey data are recalculated utilising Glen Grey population data kindly supplied by the Magistrate of Lady Frere and Dr. Von Zeynek's published cases of oesophageal cancer. The 22 locations now show only 3 which deviate from the district norm at a significant level (where $P<0,05$). Two of these locations, 3 and 20 in the north-east, fall on the high side, and one location, 12, south-west of the hospital, falls on the low side. To show the method, Table II has

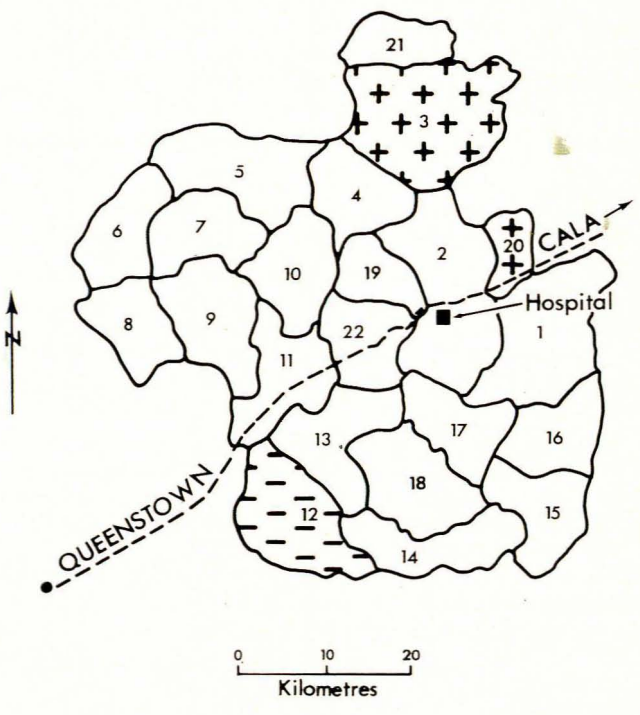

Fig. 1. Oesophageal carcinoma in the locations of Glen Grey District from 1 January 1964 to 31 March 1972. Statistically significant variations from district norm.

TABLE II. OESOPHAGEAL CANCER BY LOCATIONS (8¼ YEARS) IN THE GLEN GREY DISTRICT

(1) Location	(2) Observed cases 8¼ years	(3) Population 1972	(4) Crude incidence p.a. per 100 000	(5) Expected cases (8¼ years)	(6) Significance level (*P*)
1	12	8 246	18	8,6	—
2 (a) and (b)	11	12 198	11	12,6	—
3	15	7 182	25	7,5	0,05 (high)
4	2	3 820	6	4,0	—
5	6	5 016	15	5,2	—
6	1	2 994	4	3,1	—
7	4	4 554	11	4,7	—
8	2	2 051	12	2,1	—
9	5	4 815	13	5,0	—
10	1	2 857	4	3,0	—
11	4	4 807	10	5,0	—
12	5	12 922	5	13,5	0,05 (low)
13	5	4 110	15	4,3	—
14	6	5 090	14	5,3	—
15	4	5 065	10	5,3	—
16	3	1 821	20	1,9	—
17	5	3 092	20	3,2	—
18	1	4 134	3	4,3	—
19	3	1 925	19	2,0	—
20	5	1 059	58	1,1	0,05 (high)
21	3	2 630	14	2,7	—
22	8	6 274	15	6,5	—
Glen Grey	111	106 660	12,61	110,9	

Col. 2 supplied by courtesy of Dr Von Zeynek and col. 3 by the magistrate, Lady Frere. (Ref. 8/3 of 4 July 1972 and 27 Dec. 1973) (and as used by Dr Von Zeynek pers. comm. 7 June 1973).

the Glen Grey data with two additional columns which make allowance for probability. Column 6 shows the result of comparing the cases actually observed (column 2) with those hypothetically expected (column 5), utilising Poisson tables.[7] The resulting map (Fig. 1) shows a contrast with Von Zeynek's distribution. Location 12 becomes 'significantly low ($P<0,05$)' and is distinguished from locations 6, 10 and 18 which had previously been uniformly classed with it as 'below 5 (incidence rate) per 100 000 per annum.' Location 20, which had appeared in a class alone as 'above 40 (incidence rate)', and far above all other locational incidence rates, is now shown to be significantly high, and in a similar class to location 3, previously shown to have an incidence (20 - 29 per 100 000) only slightly above Glen Grey's over-all rate. The district annual incidence figure (12,61) is used since the purpose is to ascertain *within the district* which locations differ significantly above or below the district rate. Thus it can be seen that, far from the rates for oesophageal carcinoma in the Glen Grey district differing widely between locations, most do *not* in fact vary significantly from the district norm. This map based on probability locates the three areas whose case numbers differ at significant levels from those of the district. Their distribution at least suggests as a possibility that patients near the Mission hospital more readily attend for treatment than persons from far afield.

It must be emphasised that Choynowski's method applied here to South African material is making allowance for varying sizes of population, while at the same time removing the worst effects of random variation. Moreover, few divergences reaching significant levels are found between cases observed and those expected at location scale in the Ciskei or Transkei in oesophageal carcinoma studies generally. Thus a more suitable spatial unit for geographical study is undoubtedly the inter-district scale.

Finally, it should be noted that, while case numbers of oesophageal cancer are insufficient for geographical analysis at locational scale, that unit of area may be perfectly adequate for analysing a commoner complaint. For instance, ischaemic heart disease, claiming some 32% of all deaths in Tasmania, has recently been analysed[11] for significance at scales of area with populations-at-risk down to under a thousand in each and yet permitting a high degree of locational exactness.

For epidemiological and aetiological studies each bodily site of cancer must be treated as a separate condition, and hence absolute case numbers cannot be expected to be large. To gain an increase of absolute numbers aggregation is often employed. This too has disadvantages. Aggregation[1] over time may mask temporal changes of occurrence, and aggregation by geographical area tends to lose spatial exactness. Yet for studies at district level among the Blacks (whose causes of death are in general not yet medically certified) a wide degree of team-work between practising doctors over an area is prerequisite. In spite of Oettlé's pleas for co-operation

12 years ago, it is still a national disgrace that there is no South African national cancer registry. Had it been set up then, there would by now have been a well-documented basis for studies of the spatial variations of cancer which are well known to occur among the country's many ethnic groups. In this way, not only South Africa but world medical science, and hence Man himself, would have stood to gain. Even now, if benefit is ever to arise through sure knowledge of cancer patterns in South Africa, a national registry is essential.

This study was carried out at the Cancer Research Unit of the National Cancer Association of South Africa, Johannesburg, during sabbatical leave.

REFERENCES

1. McGlashan, N. D. *in* Dawson, J. A. and Doornkamp, J. C. eds (1973): *Evaluating the Human Environment*. London: Edward Arnold.
2. Snow, J. (1855): *On the Mode of Communication of Cholera*, 2nd ed. London.
3. Choynowski, M. (1959): J. Amer. Statl. Assoc., **54**, 385.
4. Oettlé, A. G. (1967): National Cancer Institute Monograph, **25**, 111.
5. Harington, J. S. and McGlashan, N. D. *in* Saunders, S. J. and Terblanche, J., eds (1973): *Liver*, (Proceedings of an International Liver Conference, Cape Town, January 1973). London: Pitman Medical.
6. McGlashan, N. D. (1973): *Medical Geography Techniques and Field Studies*, p. 336. Methuen: London.
7. *In* Diem, K. and Lentner, C. eds (1970): *Documenta Geigy Scientific Tables*. 7th ed., pp. 107 - 108. Basle: Geigy.
8. Burrell, R. J. W. (1969): J. Nat. Cancer Inst., **43**, 877.
9. Von Zeynek, E. R. (1973): S. Afr. Med. J., **47**, 325.
10. *Idem* (1973): *Ibid.*, **47**, 329.
11. McGlashan, N. D. (1972): Unpublished paper of the International Geographical Union Symposium of Medical Geography, Guelph, Canada, August 1972.