able calculator. NIH Publication 79–1649, US Government Printing Office, Washington DC 1979.
17. Harvey EB, Brinton LA. Second cancer following cancer of the breast in Connecticut, 1935–1982. *Natl Cancer Inst Monogr* 1985, **68**, 99–112.
18. Storm HH, Jensen OM. Risk of contralateral breast cancer in Denmark 1943–1980. *Br J Cancer* 1986, **54**, 483–492.
19. Murakami R, Hiyama T, Hanai A, Fujimoto I. Second primary cancers following female breast cancer in Osaka, Japan—a population-based cohort study. *Jpn J Clin Oncol* 1987, **17**, 293–302.
20. Schwartz AG, Ragheb NE, Swanson GM, Satariano WA. Racial and age differences in multiple primary cancers after breast cancer: a population-based analysis. *Breast Cancer Res Treat* 1989, **14**, 245–254.
21. Ewertz M, Storm HH. Multiple primary cancers of the breast, endometrium and ovary. *Eur J Cancer Clin Oncol* 1989, **25**, 1927–1932.
22. Horn PL, Thompson WD. Risk of contralateral breast cancer: associations with histologic, clinical and therapeutic factors. *Cancer* 1988, **62**, 412–424.
23. Lavey RS, Eby NL, Prosnitz LR. Impact of radiation therapy and/or chemotherapy on the risk for a second malignancy after breast cancer. *Cancer* 1990, **66**, 874–881.
24. Schwartz AG, King MC, Belle SH, Satariano WA, Swanson GM. Risk of breast cancer to relatives of young breast cancer patients. *J Natl Cancer Inst* 1985, **75**, 665–668.
25. Horn PL, Thompson WD. Risk of contralateral breast cancer: associations with factors related to initial breast cancer. *Am J Epidemiol* 1988, **128**, 309–323.
26. Prior P, Waterhouse JAH. Multiple primary cancers of the breast and ovary. *Br J Cancer* 1981, **44**, 628–636.
27. Ewertz M, Mouridsen HT. Second cancer following cancer of the female breast in Denmark, 1943–1980. *Natl Cancer Inst Monogr* 1985, **68**, 325–329.
28. Kelsey JL, Hildreth NG. *Breast and Gynecological Cancer Epidemiology*. Boca Raton, Florida, CRC Press, 1985.
29. Brinton LA, Fraumeni JF. Epidemiology of uterine cervical cancer. *J Chron Dis* 1986, **39**, 1051–1065.
30. Senofsky GM, Wanebo HJ, Wilhelm MC, *et al.* Has monitoring of the contralateral breast improved the prognosis in patients treated for primary breast cancer? *Cancer* 1986, **57**, 597–602.

Eur J Cancer, Vol. 29A, No. 10, pp. 1414–1418, 1993.
Printed in Great Britain

0964–1947/93 $6.00 + 0.00

Response of a Cancer Registry to Reports of Disease Clusters

Will D. King, Gerarda A. Darlington, Nancy Kreiger and Gordon Fehringer

A protocol has been developed to investigate and report perceived clusters of cancer using a population-based cancer registry. The protocol comprises a series of steps which lead to assessment of the cluster's importance on the basis of three criteria: (1) statistical evidence of clustering; (2) documentation of the existence of exposure to a carcinogen; and (3) biological plausibility of the relationship between the exposure and the cancer of interest. The evaluation of these criteria results in one of three recommendations: further study, surveillance only, or no action. The protocol provides a systematic approach for investigation, makes efficient use of available cancer registry data, and responds to public concerns. The protocol is demonstrated by its application to an inquiry concerning an apparent excess of lung cancer in a small Ontario town and the possible role of radon gas exposure. The public health importance and limitations of addressing perceived disease clusters are discussed.
Eur J Cancer, Vol. 29A, No. 10, pp. 1414–1418, 1993.

INTRODUCTION

In 1990, the Ontario Cancer Treatment and Research Foundation (OCTRF) received over 150 inquiries concerning the occurrence of cancer. Most of these were addressed with existing cancer statistics derived from the Ontario Cancer Registry (OCR); a number, however, concerned potential cancer clusters, a type of inquiry for which no formal response guidelines existed. As a result, a protocol was developed to investigate reports of perceived cancer clusters using registry data. The investigation of such reports provides an important public service and has the potential to identify environmental risk factors for cancer. This paper describes the protocol which has been developed and discusses its application to a recent citizen concern.

Correspondence to W.D. King.
The authors are at the Division of Epidemiology and Statistics, The Ontario Cancer Treatment and Research Foundation, 620 University Avenue, Toronto, Ontario, Canada M5G 2L7.
Revised and accepted 9 Mar. 1993.

SUBJECTS AND METHODS

The development of the cancer cluster protocol was based on existing informal procedures of the OCTRF, a review of protocols used by other disease control agencies [1–4], and consideration of available resources. This protocol is distinguished from others by its use of incidence data derived from a cancer registry rather than the informant. The OCR contains records on residents of Ontario diagnosed as having cancer since 1964. This resource permits us to carry out temporal and geographical investigations of cancer incidence in an expeditious and non-invasive manner.

Protocol

The cluster protocol is applied to those inquiries which concern associations among cancer cases characterised by: (i) a defined geographical area and/or time period, (ii) a localised environmental exposure, (iii) an institutional exposure (e.g. hospital, school), or (iv) an occupation. The protocol comprises six steps, as follows.

Step 1: *preliminary assessment and communication*

The inquiry is assessed in terms of the specificity of cancer type, the sufficiency of the number of cases, and the plausibility of the association with a suspected exposure; it is continued if any of these conditions are met. To satisfy specificity of cancer type, all cases should be of one type or body system. Many reported clusters are a heterogenous collection of malignant tumours for which a common cause is unlikely. Small numbers are often a problem in cluster investigations. Typically, a meaningful analysis of cancer rates according to available geographical units cannot be performed with less than 5 cases [5], unless the site or histological type of cancer is extremely rare. The plausibility of the suspected exposure is evaluated as to its merit for further consideration in the investigation. Often exposures suggested by the informant can be ruled out based on the temporal relationship between exposure and case diagnosis.

Informants are notified of the decision to continue, the specific issues which may arise in the particular investigation, and the possible outcomes. When an investigation is discontinued, informants are provided with an explanation of this decision, an informational brochure discussing cancer clusters, and specific information on cancer rates and known risk factors which may help them to understand what they have observed.

Step 2: *problem identification and verification of cases*

Complete problem identification takes place as an investigation continues. Spatial, temporal or other relevant boundaries are determined through communication with informants and consultation with an epidemiologist. Incident cases are identified in the OCR. Recently diagnosed cases or other valid cases identified by the informant and not identified in the OCR, may be verified by hospital and pathology records. Verification of diagnosis and residence classification, by means of a review of all records for the cases of interest, are undertaken to minimise reporting bias.

Step 3: *analysis*

Statistical analysis is used to describe the reported cluster and to determine statistical significance. Geographical analysis is based on census subdivisions, or on more specific designations if required.

Firstly, age- and sex-standardised incidence rates (SIR) for the cancer sites and geographical areas of concern are compared with those of Ontario and census divisions of Ontario. Next, statistical methods designed for the study of spatial, temporal or space–time disease clusters are applied. Overall statistical importance is indicated by the amount of evidence for a disease cluster, provided by one or more of the appropriate statistical tests. Because of the *post hoc* nature of the analysis, the cluster statistics employed are not intended to test a scientific hypothesis. These tests should be used in a descriptive manner to evaluate the importance of the reported cluster.

When investigating cancer clusters, the identification of spatial clusters is usually of most concern [6]. However, in some circumstances it may be appropriate to consider temporal or space–time clusters. It is recognised that a large body of literature exists regarding tests for clusters of events. Several specific tests were selected based on the type of data commonly available for these investigations and based on issues of computational and statistical efficiency. In the OCTRF protocol, methods commonly used for detecting spatial clusters include the Whittemore statistic [6], the Ohno statistic [7] and the Geary ratio [8]. The methods used for detecting temporal clusters are the SCAN statistic [9, 10] and the approach of Ederer, Myers and Mantel [11]. For detecting space–time clusters, the method of Knox [12] and the approach of Barton and David [13] are applied. More complex graphical procedures exist for identifying clusters [14–16]. These methods are appropriate if more detailed study is required, but are not used within the framework of this cluster protocol.

Step 4: *assessment of biological plausibility and documentation of suspected exposure*

Determination of biological plausibility of the hypothesis is based on aetiological similarities among the types of cancer involved, plausibility of the association with the suspected exposure, and sufficiency of the latent period between exposure and onset of cancer. Previous epidemiological, experimental and animal studies can be used to evaluate the biological plausibility by indicating mechanisms of disease or by corroborating the basic association between exposure and disease. This criterion should be used cautiously since it could obstruct the investigation of previously undetected causes of cancer. Documentation of the existence and location of potential exposures is done through contact with the Ontario Ministries of Health, Labour and the Environment, as well as other agencies.

Step 5: *determination of cluster importance*

The importance of the cluster and subsequent action to be taken are determined by the statistical importance, biological plausibility and documentation of exposure. Statistical importance is satisfied if a statistically significant association is observed for any of the appropriate tests. Biological plausibility is subjectively assessed by reviewing the information gathered in step 4 and through consultation with appropriate experts. Likewise, an objective evaluation of the exposure may not be possible and a subjective assessment may have to be made based on the available data.

One of the following recommendations is made: (i) no further action, (ii) future surveillance, or (iii) detailed study. If only one of the three criteria is met, then no further action is recommended. If both biological plausibility and statistical importance are established, further study is necessary. Otherwise, monitoring of rates is recommended.

Step 6: *preparation of final report*

A final report, containing the recommendations, is sent to the informants, with copies possibly sent to the local Medical Officer of Health and the OCTRF files.

RESULTS

The following example demonstrates the application of the protocol to a recent inquiry. A citizen contacted the OCTRF with a concern over the number of lung cancers he had observed in his town in the last 5 years.

Step 1: *preliminary assessment and communication*

Seventy-four incident cases of lung cancer were diagnosed between 1984 and 1988 in town residents. Radon gas in this area was identified as a possible cause; a plausible association between domestic radon exposure and lung cancer has been hypothesised [17–19].

Step 2: *problem identification and verification of cases*

Two definitions of the study area were created to explore clustering of different spatial magnitudes. The first (the Town)

was the census subdivision, representing the town of interest which had a population of about 9000 in 1986. This designation was used to make comparisons with Ontario rates and rates in other similar towns, and to explore clustering of lung cancer cases within the town.

The second study area (the Area), comprising the 33 census subdivisions within 75 km of the town, was used to investigate clustering according to larger spatial boundaries. The Area included five towns, four villages, and 24 townships. In 1986, the total population for the area was 95 000 and the largest town had a population of 15 000.

From the OCR, 334 cases were identified in the Area and 74 cases within the Town, for the years 1984 to 1988. Because of the nature of OCR source records, there existed the potential for reporting error regarding usual residence. In particular, there is a tendency for cases in rural settings to report the nearest population centre as their residence. Such bias would result in the appearance of elevated incidence rates in census subdivisions defining population centres. In order to rule out this possibility, residence verification was performed. As a result, 2 cases were removed from, and 20 cases changed census subdivisions within the Area. Of the 74 cases initially identified in the Town, 15 cases were resolved to residences outside the Town. The residences of 3 cases from the Area were changed to the Town, for a total of 62 cases in the Town.

Residence verification was not performed for lung cancer cases with residence codes outside of the Area. Thus, there is a potential for incidence rates of census subdivisions at the borders of the Area to be biased according to the pattern of residence misclassification described above.

Step 3: *analysis*

A significant excess in lung cancer incidence in the Town was observed for both males and females as compared to Ontario rates, prior to residence verification. SIR estimates in comparison with the Ontario population were changed somewhat following residence verification (see Table 1). While the two-fold excess in lung cancer was still observed for males, the residence-corrected SIR for females indicated no excess.

A comparison with similar Ontario towns was performed in order to discern whether the elevation in rates could be explained by the comparison population used. Rates prior to residence verification were employed to make the groups similar with respect to residence misclassification. Eleven towns were selected to form a comparison population according to population

Table 1. *Standardised incidence ratio (SIR) estimates and* 95% *confidence intervals (CI) for incidence rates in the Town*

Comparison population	Sex	SIR	(95% CI)
All Ontario			
Prior to residence verification	Male	2.4	(1.8, 3.1)
	Female	1.7	(1.0, 2.5)
After residence verification	Male	2.2	(1.6, 2.9)
	Female	1.1	(0.6, 1.8)
Similar Ontario towns*			
Prior to residence verification	Male	1.7	(1.2, 2.2)
	Female	1.2	(0.7, 1.8)

*Eleven towns similar in terms of population, rural setting and census characteristics of age, land area, ethnicity, education, household income, and industry.

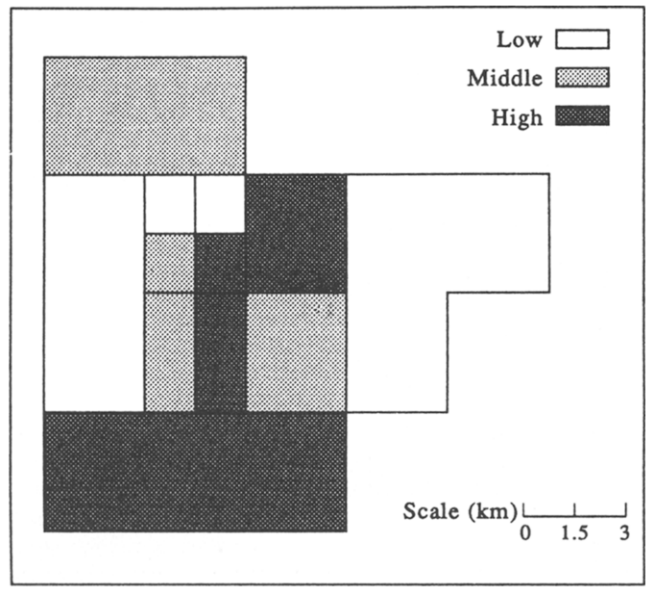

Fig. 1. Map of approximate incidence rates (both sexes combined) for subregions of the Town. The distribution of rates was divided into thirds to form low, middle and high subregions.

and rural setting and census characteristics of age, land area, ethnicity, education, household income and industry [20, 21]. Lung cancer rates in the Town were significantly elevated for males [SIR = 1.7, 95% confidence interval (CI) = 1.2–2.2] and were not significantly elevated for females (SIR = 1.2, 95% CI = 0.7–1.8) in this comparison. The estimates of excess for both sexes were smaller than observed in the comparison with Ontario rates, suggesting that some of the observed excess may be due to risk factor characteristics or reporting biases common to this type of town.

Spatial cluster methods were used here, as there was no suggestion of a temporal effect. The Whittemore test statistic [6] is based on the mean distance between all pairs of cases, where each case is assumed to reside in the centre of a subregion. The Geary statistic [8] examines variation in cancer rates between subregions and integrates that with information on the spatial structure of the area. The Ohno test [7] is similar to Geary using rate categories instead of interval data.

In order to compute the spatial statistics, the Town was divided into 12 subregions. For each subregion, the frequency of lung cancer and an approximate population figure were obtained. To approximate population, the number of residences was counted and assumed to be proportional to the population. Figure 1 contains a plot of the approximated incidence rates for each of the subregions. The Whittemore method resulted in a test statistic of marginal significance (P value = 0.07), the Ohno test had a higher P value (0.15), and the Geary ratio, for which values close to 0 indicate clustering and values close to 1 indicate no clustering, revealed no evidence of clustering (see Table 2).

Table 2. *Application of cluster statistics to the Town of interest and to the Area*

Test statistic	Town statistic (P value)	Area statistic (P value)
Whittemore (Z)	−1.80 (0.07)	−0.66 (0.26)
Ohno ($\chi^2_{df=1}$)	1.78 (0.15)	1.13 (0.24)
Geary ratio (GR)	0.93 (0.31)	0.99 (0.47)

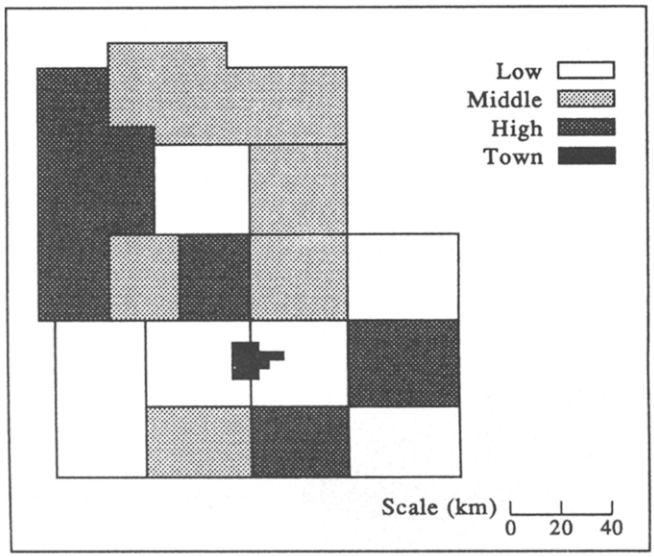

Fig. 2. Map of age-adjusted incidence rates (both sexes combined) for subregions of the Area. The distribution of rates was divided into thirds to form low, middle and high subregions.

The Area was divided into 16 subregions, one of which was the Town (Fig. 2). No obvious clustering of rates was observed on visual inspection. The results of performing the statistical tests are presented in Table 2; no evidence of clustering in the Area was observed using any of the three methods.

Step 4: *assessment of biological plausibility and documentation of exposure*

Radon, the exposure of concern, is a by-product of uranium decay and a naturally occurring radioactive gas. There is epidemiological evidence that exposure to high doses of radon is associated with lung cancer, although the risk associated with lower domestic doses is unclear [17–19].

While no organised survey of radon levels had been conducted in the Area, the available evidence indicated that high levels are plausible. High levels (5–10 pCi/l) were reported in a survey 50–75 km from the Town [22] and measurements available for homes in the Town [23] were above the Canadian average, with the highest being at the Canadian action level.

Step 5: *determination of cluster importance*

The established association between lung cancer and radon gas makes the cluster biologically plausible, although elevation in risk associated with exposure in homes is not established. There is incomplete but indicative evidence of the existence of high levels of radon. Finally, overall there was no statistical evidence of clustering. However, SIR for the Town remained statistically significant after adjustment for residence misclassification and in comparisons with other similar towns. The decision was made, therefore, to monitor lung cancer rates in the Area over the next 3 years.

Step 6: *preparation of final report*

The final recommendation for future surveillance was described to the informant, and a procedure for monitoring the rates was established.

DISCUSSION

Several agencies have reported their methods for the investigation of reported cancer clusters [1–4]. The OCTRF protocol has incorporated many of the strategies common to these protocols, such as early communication with the informant, methods of case ascertainment, statistical analysis, prioritising reports for in-depth study, and the reporting of results. The uniqueness of this protocol is its utilisation of, and reliance on, the resources available in a population-based cancer registry. All cancer information is obtained from the registry, rather than the informant; this ensures uniformity of diagnosis, completeness of ascertainment, and ease of calculation of expected frequencies.

The example presented illustrates the actions taken in response to a specific cluster inquiry. Reports of apparent cancer clusters received by cancer registries may vary immensely in the scale of the problem and nature of the suspected risk factor. However, we suggest applying the same general steps to employing registry resources in the investigation of each inquiry.

The protocol provides an essential response to inquiries concerning cancer clusters using the resources of a cancer registry; there are, however, several limitations to this registry-based approach which should be acknowledged.

Analyses of clusters are often restricted to census subdivisions, for which disease frequencies are sufficient for meaningful analysis, and for which the populations at risk are available. Of interest in this investigation was the cancer distribution within the Town. To this end, *ad hoc* estimates of the population distribution were made but could not include the underlying age and sex distribution.

The cluster protocol relies to a large extent on the OCR, which has an acknowledged degree of error with respect to reporting completeness, diagnostic accuracy, and residence coding [24]. These could affect cluster analysis if there is a differential error according to any of the units (e.g. time, space, disease) under study. Verification of diagnosis and residence, as was performed here, is not always possible.

The protocol is also limited in its capacity to identify or confirm environmental contaminants and exposures. The investigation presented relied on incomplete geological evidence of environmental exposure.

In general, the statistical analysis of data from perceived disease clusters has been cautioned on several grounds, including the concealed presence of multiple comparisons and *post hoc* hypothesis generation [3, 5, 25]. It is acknowledged that this type of criticism is valid and that results must be interpreted carefully.

Several benefits arise, however, with the use of this cluster protocol. First is the provision of a systematic approach, facilitating an efficient investigation without a major commitment of resources. Second is the potential for generating hypotheses concerning environmental causes of cancer. Third is extending the utility of a population-based cancer registry. Finally, and most importantly, the protocol fulfils a public service, by facilitating a response to community concerns and by educating the public on the occurrence of cancer.

1. Fiore BJ, Hanrahan LP, Anderson HA. Public health response to reports of clusters. State health department response to disease cluster reports: a protocol for investigation. *Am J Epidemiol* 1990, **132** (Suppl. 1), S14–S22.
2. Devier JR, Brownson RC, Bagby JR, *et al.* A public health response to cancer clusters in Missouri. *Am J Epidemiol* 1990, **132** (Suppl. 1), S23–S31.
3. Bender AP, Williams AN, Johnson RA, Jagger HG. Appropriate public health responses to clusters: the art of being responsibly responsive. *Am J Epidemiol* 1990, **132** (Suppl. 1), S48–S52.
4. Centers for Disease Control. Guidelines for investigating clusters of

health events. *Morbidity and Mortality Weekly Rep*, Atlanta, 1990, **39** (RR-11).
5. Neutra RR. Counterpoint from a cluster buster. *Am J Epidemiol* 1990, **132**, 1–8.
6. Whittemore AS, Friend N, Brown BW, Holly EA. A test to detect clusters of disease. *Biometrika* 1987, **74**, 631–635.
7. Ohno Y, Aoki K, Aoki N. A test of significance for geographic clusters of disease. *Int J Epidemiol* 1979, **8**, 273–281.
8. Geary R. The contiguity ratio and statistical mapping. *Incorporated Statist* 1954, **5**, 115–145.
9. Weinstock MA. A generalised scan statistic test for the detection of clusters. *Int J Epidemiol* 1981, **10**, 289–293.
10. Wallenstein S. A test for the detection of clustering over time. *Am J Epidemiol* 1980, **111**, 367–372.
11. Ederer F, Myers MH, Mantel N. A statistical problem in space and time: do leukemia cases come in clusters? *Biometrics* 1966, **20**, 626–638.
12. Knox G. The detection of space-time interactions. *Appl Statist* 1964, **13**, 25–29.
13. Barton DE, David FN. The random intersection of two graphs. In David FN, ed. *Research Papers in Statistics*. New York, John Wiley & Sons Inc, 1966, 455–559.
14. Selvin S, Shaw G, Schulman J, Merrill DW. Spatial distribution of disease: three case studies. *JNCI* 1987, **79**, 417–423.
15. Selvin S, Merrill D, Schulman J, *et al.* Transformations of maps to investigate clusters of disease. *Soc Sci Med* 1988, **26**, 215–221.
16. Openshaw S, Craft AW, Charlton M, Birch JM. Investigation of leukaemia clusters by use of a geographical analysis machine. *Lancet* 1988, **i**, 272–273.
17. Samet JM. Radon and lung cancer. *JNCI* 1989, **81**, 745–757.
18. Harley NH, Harley JH. Potential lung cancer risk from indoor radon exposure. *A Cancer J for Clinicians* 1990, **40**, 265–275.
19. Bowie C, Bowie SHU. Radon and Health. *Lancet* 1991, **337**, 409–413.
20. Canada Census 1986: *Census Divisions and Subdivisions; Ontario: Part 1—Profiles*. Cat. No. 94-111, Ottawa, Statistics Canada, 1987.
21. Canada Census 1986. *Census Divisions and Subdivisions; Ontario: Part 2—Profiles*. Cat. No. 94-112. Ottawa, Statistics Canada, 1987.
22. Radiation Protection Bureau. Department of National Health and Welfare (personal communication, Feb 1991).
23. Alphanuclear Company, Missisauga Ontario (personal communication, Feb 1991).
24. Clarke EA, Marrett LD, Kreiger N. *Twenty Years of Cancer Incidence, 1964–1983: The Ontario Cancer Registry*. Toronto, The Ontario Cancer Treatment and Research Foundation, 1987.
25. Rothman KJ. A sobering start for the cluster busters' conference. *Am J Epidemiol* 1990, **132** (Suppl. 1), S6–S13.

Acknowledgements—This study was supported by the Ontario Cancer Treatment and Research Foundation.

Eur J Cancer, Vol. 29A, No. 10, pp. 1418–1423, 1993.
Printed in Great Britain

0964–1947/93 $6.00 + 0.00

Cancer Risks Among European Migrants in Sáo Paulo, Brazil

Christine Bouchardy, Myriam Khlat, Antonio Pedro Mirra and Donald Maxwell Parkin

Using both mortality and incidence data, cancer risk in Italian, Spanish and Portuguese migrants to Sáo Paulo were compared with those in the Brazil-born population, and with those in their countries of origin. Italian and Spanish migrants show changes in cancer risks which are rather similar to those observed in migrants of the same origin in other parts of South America: they increase their rates of oropharyngeal, oesophageal, cervical and breast cancers and they decrease their rates of lung cancers. However, for cancer of the oesophagus, the changes are greater in Sáo Paulo, where migrants acquire rates similar to those of the natives. For colon cancer, rates in Italian migrants decrease in the low risk area of Sáo Paulo and increase in the high risk area of Argentina. Changes in Portuguese migrants are less evident: their rates of colorectal cancer remain high, and, as found for Japanese migrants in Sáo Paulo, they also retain their higher risks of stomach cancer.
Eur J Cancer, Vol. 29A, No. 10, pp. 1418–1423, 1993.

INTRODUCTION

STUDIES OF cancer risk in European populations migrating to South America are scarce. Death rates from cancer among migrants in Uruguay and Argentina have recently been published, and show interesting patterns, especially for digestive cancers [1, 2]. The only published studies of migrants to Brazil concern Japanese in Sáo Paulo [3–5], although the cancer pattern in migrants to Brazil is of particular interest, as there are unusually high local rates of oral cavity, oesophageal, stomach and cervix uteri cancers.

The European migration to Brazil dates back to the beginning of the nineteenth century, due in part to the development of extensive agriculture, mainly coffee and cotton, and to the existence of large unpopulated areas in the southern part of the country. This movement was slowed down by the Brazilian government at the time of the coffee crisis in 1930, and the only county which had enough financial resources to maintain the immigration level was that of Sáo Paulo. In Sáo Paulo, the main

Correspondence to D.M. Parkin.
D.M. Parkin is at the International Agency for Research on Cancer, 150 Cours Albert Thomas, 69372, Lyon Cedex 08, France; C. Bouchardy is at the Geneva Tumour Registry, Switzerland; M. Khlat is at the Institut National d'Etudes Démographiques, Paris, France; and A.P. Mirra is at the Sáo Paulo Tumor Registry, Brazil.
Received 9 Nov. 1992; accepted 8 Feb. 1993.