

# Registration of cancer in Flanders as a means of determining environmental health effects

by

Van Herbruggen J. <sup>1</sup>, Michielsen A. <sup>1</sup>, Wouters K. <sup>1</sup>,  
Droste J. <sup>1</sup>, van Sprundel M. <sup>1</sup>, Weyler J. <sup>1</sup>

---

## Abstract

**Purpose of the study:** *One of the objectives of the registration of cancer is to represent the geographical distribution of the incidence of cancer, which, in combination with other information, could be used to evaluate environmental health effects. This study examines whether a cancer registry could be used to evaluate health effects of point sources, and if so, what the necessary requirements of registration are, and to what extent a (cancer)registry needs to contain geographical information in order to evaluate environmental health effects in a meaningful way.*

**Main findings:** *To be useful for the investigation of a cancer (disease) cluster, a registry needs to fulfil several conditions, e.g. it should be as complete and as accurate as possible and duplicated data should be avoided. The rapid availability of data is very important. In Flanders, a network of the existing cancer registries with linking of data can represent an important added value for the study of environmental health effects through cancer registration. When a suspected cluster is reported, it*

---

Correspondence: van Sprundel M., University of Antwerp, Epidemiology and Community Medicine, Universiteitsplein 1, 2610 Antwerp, Belgium.

<sup>1</sup> Epidemiology and Community Medicine, University of Antwerp.

*often represents a very local increase in incidence of disease. For this reason, the registry needs to contain detailed geographical information. The most appropriate level of registration for the study of a cluster seems to be the statistical code. The inventory of the incident cancer cases on the level of statistical units is possible, but very labour-intensive because of the current privacy-legislation.*

**Principal conclusions:** *Cancer registration can be used to study the geographical distribution of cancer incidence in relation to environmental exposure to health risk factors. If the authorities consider this to be an important objective of a cancer registry, then a legislation on the registration of cancer is necessary. This legislation should lead to a more efficient organisation of cancer prevention by means of a well-structured cancer registry compiled in an ethically, economically, and scientifically acceptable manner.*

## **Keywords**

Cancer registration, environmental health, cluster analysis.

## **Introduction**

In Flanders, several initiatives have been undertaken with regard to the registration of the incidence of cancer. In a cancer registry a number of previously determined data concerning newly diagnosed cancer cases are collected and registered. The principal aim of these registries is to produce statistics on cancer incidence in well-defined population groups and to generate the basis for descriptive and analytic epidemiological research. One of the objectives of the registration is to represent the geographical distribution of the incidence of cancer, which, in combination with other information, could be used to evaluate environmental health effects. This study aims to answer the question of whether a cancer registry could be used to evaluate health effects of point sources of pollution and if so, what the necessary requirements, including geographical information, are.

## **Methods**

In the national and international literature, a search for information concerning cluster investigation in relation to environmental pollution, the

use of cancer registration for this kind of investigation and the different geographical units and their application was set up. The Medline database was consulted (using the search terms “cancer” and “registry” and “environmental health” (1966-2000), and “cancer” and “registry” and “cluster” (1966-2000)), as well as handbooks and references in handbooks. In addition, conditions regarding the geographical identification of the cancer registration in Flanders required for the investigation of possible health effects caused by polluting point sources were drawn up. Finally, the performance of the different registration systems in Flanders with respect to the geographical identification of the registered cases was examined.

## Results

### 1. Cluster investigation

#### a) Introduction

In general, the study of the geographical distribution of cancer in order to evaluate environmental health effects can start from two different points. On the one hand, it can be a routine action to send alerts in the case of raised incidence. On the other hand, the study can be initiated out of concern for the health effects of a known polluting point source. An increased cancer incidence around a polluting point source is frequently reported (1). The reporters are often the local inhabitants (e.g. people living near a waste incinerator), however other sources (e.g. a local general practitioner) may also report the suspicion of a higher cancer incidence around a polluting source. The alert can be made public by the different media (e.g. (local) newspapers, radio, television, ...). By doing so, the reporters exert pressure on the (local) authorities to force them to evaluate the local situation, give feedback to the local inhabitants and the media, and if necessary to take appropriate action. This evaluation task is often delegated to researchers in the field of environmental health. The (scientific) approach of such an announcement should be an integrated one. The collaboration of several scientific disciplines is required (e.g. epidemiology, toxicology and statistics). Recently, a separate section, small area epidemiology, has been developed from these disciplines, which focuses on the approach to this issue.

#### b) Principles

In most cases, researchers go through a step by step analysis, with a disease-track, an environment-exposition-track and some communica-

tion aspects (1). In each phase, communication is extremely important. The researchers have to communicate with and to report frequently to the authorities, which commissioned the investigation of the disease cluster. Furthermore, there has to be communication with the local inhabitants and pressure groups (e.g. local anti-pollution groups), who have pointed out the problem.

In most cases, the analysis of a supposed disease cluster around a polluting point source will start with a rough testing of the disease-cluster-signal and the environment-exposition-cluster signal (1). The disease-cluster-signal will be verified on the basis of the available data on incidence, prevalence and mortality (mostly rough figures, e.g. national mean). The environment-exposition-cluster-signal will be roughly verified by carrying out an on-site inspection to obtain an impression of the most relevant polluting sources and the possibilities of exposure. If both disease-cluster-signal and environment-exposition-signal are not refuted after the rough testing, the relationship between the two will be roughly reviewed on the basis of a literature study. The central question here is: "Can the assumed environmental pollutant be causally associated with the disease concerned?" If the exposition-disease relationship is not refuted at this rough testing stage, further research is justified.

To arrive at a conclusion as to whether a cluster exists or not, the incidence of the disease in the potentially exposed population (index population) has to be compared with the incidence of the disease in a reference population. Ideally, the reference population should be comparable to the index population in all aspects (except for the suspected environmental factor).

When a cluster is analysed, the index population is defined in time and space, after which the number of observed disease cases in the index population will be determined and compared with the number of expected disease cases. The latter can be calculated by using available data concerning the disease under study in a reference population, after correction for age and sex. This implies the need for age- and gender specific data regarding the incidence of disease in the reference population. Furthermore, demographic data (number of inhabitants and their gender and age-distribution) from the index- and reference populations of the period under investigation are also required.

Initially, the statistical analysis of clusters was limited to the level of the geographical units for which disease- and population data were available. Recently, methods have been developed to analyse the incidence

of disease around a point source (2, 3, 4, 5, 6, 7). Often, these methods are based on the analysis of the distance between pairs of cases. So, the NNA-test (Nearest Neighbour Areas) inspects each case separately and determines whether it is unusually close to its nearest neighbours (3). If so, it is said to be a “clustered case” (3, 5). A calculation is then made of how many of these cases could be expected if the distribution was random. The Monte Carlo method is used to investigate whether the observed number of clustered cases significantly exceeds the expected number of cases.

For the Cuzick-Edwards test (6), control locations, which are generated by computer algorithms, are used to compare the spatial distribution of the locations of the cases with the spatial distribution of the locations of the controls. Another method (8) is not based on distances, but compares the distribution of the number of cases in population units of approximately the same magnitude. These population units are composed of small administrative units for which population data are known.

Another way to investigate clustering in time and place is the “all possible pairs method” (9, 10) The cases are registered in place and time, then any two cases are evaluated for being a pair of cases by means of a previously determined definition. The next step is to evaluate whether these pairs of cases have an unusual tendency to cluster in time and place. If there are significantly more pairs of cases than could be expected based on chance it is concluded that the clustering is unusual.

### *c) Problems*

The main disadvantage of these and other methods (11) is the fact that the “nearness” in time and place is determined arbitrarily.

In most cases, the exposed population has a limited size. Consequently the number of patients (index population) is very small. This causes some problems in interpreting the data. Because of this small index population, it is very difficult to obtain a statistically significant difference between the number of expected and observed cases. When, on the other hand, the limits of the population are less strictly defined, the potential effect becomes diluted and a statistically significant difference may not be demonstrated. In some cases, this problem can be solved by aggregation of data including more years or different clusters around a similar point source (“pooling”).

In addition, when the expected number of cases needs to be calculated, the unavailability of morbidity data (e.g. in cancer registries) of

limited geographical units may be a problem. In Flanders, registration usually takes place on postal code level. The postal code encloses 85 to 450.000 inhabitants (Personal Communication. Nationaal Instituut voor de Statistiek). Consequently, more general statistics on disease incidence, which do not always apply to the index population from which the exposed population originates are used.

The question remains to what extent the addition of more detailed geographical characteristics to the registry of incidence of cancer can result in an additional value to cluster research and which geographical level is the most desirable.

In the past, a large number of reported supposed disease clusters were not confirmed by investigation. If this is the case, it is important to keep the public informed, so that their concerns are alleviated. Only after confirmation of the existence of the disease cluster, an attempt should be made to investigate the possibility of a causal link between exposure and environmental pollution. A rational time-relation between exposure and effect is essential. A causal relationship can be excluded if the induction and latent periods are longer than the period lived in the investigated region (12). If this is not the case, however the observed associations may still be obscured by confounding factors, e.g. socio-economic status and/or life style.

## *2. Cancer registration*

### *a) Requirements*

With regard to the addition of geographical characteristics to a cancer registry, the following considerations have been taken into account: the smaller the registration units, the larger the discriminating capacity; however, the numbers will be lower with more variation due to change. A possible solution to these problems is data aggregation. Furthermore, the smaller registration units may result in a certain loss of privacy. With regard to this problem, a balance has to be found between concerns about health and protection against a life-threatening disease and concerns of protection of privacy and autonomy.

For analyses based on registration (incidence) rates in small areas, it is necessary for population denominators to be available for the same geographical boundaries as for the observed cancer data. This will limit the detail of geography that can be analysed to the smallest units that are coded or can be formed by aggregation of coded units, in both the cancer

and population data, with age- and sex-specific information available for both. Two additional problems are the possibility that the area of concern may eventually exceed the administrative boundaries, and that the size of the administrative units may hide local effects (13). Also, administrative boundaries may change.

For small area analysis, the time-period for which data are available is very important because these studies are generally based on small populations with only few cases of cancer each year, and data from several years are therefore needed to achieve reasonable accuracy of risk estimation (in small areas).

With regard to the performance of small area analyses, a population based cancer registry is preferred over a hospital based cancer registry. Population based data are desirable if more precise estimates of risk, with less chance of bias, are to be attained, and are needed particularly if smaller differences in risk according to geography are to be investigated satisfactorily.

Ideally, the registry should reach one hundred percent completeness. In addition, the data need to be accurate and without duplication. The data set has to include all the variables required for analyses (at least one site of cancer, age, sex and geographical location). However, data on possible confounding factors, e.g. occupational exposure, are often missing. For example, an analysis of disease incidence near a point source that does not take into account socio-economic confounding may be seriously biased and therefore result in misleading or false conclusions about the possible effects of pollution on the disease incidence (14).

Furthermore, there may be alternative explanations for geographical patterns of cancer incidence data e.g. differences in presence of medical care and in diagnostic practice, late registrations, accuracy and appropriateness of population denominators and differing causal factors (e.g. smoking habits) between geographical areas (15). Random variation (chance) is also a possible explanation of any geographical pattern of cancer incidence.

It would also be useful to have complete residential histories. Usually only one address per cancer patient is registered, as most studies of clusters use, inevitably, addresses at diagnosis. In cases of chronic exposure, cancer risk should be related to the duration of residence in the area. On the other hand, geographical analyses can also be related to an exposure that occurred over a brief period of time (e.g. Hiroshima).

Because of the social dimension inherent to environmental health problems, the delay in the availability of the data is a very important factor. If a suspicion of an increased cancer incidence is reported, whether it is associated with a possible pollution source or not, it is important to refute or to confirm the existence of a disease cluster as soon as possible. A cancer registry has to be designed in such a way that the desired information can be obtained quickly and easily so that this information is as recent as possible.

*b) Situation in Flanders*

To be useful for the investigation of disease clusters, a registry needs to fulfil several conditions. When a suspicion of a cluster is reported, it often involves very local increases in disease incidence. The registry must include geographic information that is as specific as possible. Generally, in Flanders the postal code is the smallest geographical unit for which incidences can be calculated. For the investigation of clusters, this is insufficient. The best-suited unit seems to be the statistical code. In Belgium, the municipalities are divided into statistical units. Each unit has a statistical code. Theoretically, one unit encloses at least 200 inhabitants. The maximum number of inhabitants of an unit is not determined, but depends on the population density. It is the smallest geographic unit on which demographic data are available. The addition of the statistical code of residence of the patient at the time of registration may simplify the subsequent investigation of clusters. If uncoded data can be registered, the coupling to statistical sectors can be performed routinely, and this would be possible if we had a legislation on the registration of cancer.

*c) Privacy legislation*

In Flanders, the setting up of a network between the existing cancer registries with coupling of data could greatly increase the value of cancer registration, because none of these registries on his own can give an exact picture of cancer in Flanders (16).

Well-structured and well-functioning registration systems exist in the Scandinavian countries (17, 18, 19, 20, 21). These registries enhance the opportunities for epidemiological research. In Belgium, epidemiological research is more difficult because of strict privacy regulations, which limits the use of cancer registries (22). Moreover, specific legislation on cancer registration is lacking. For years, Belgium has not had any legislation concerning the protection of personal medical data. Recently, a regulation has been introduced in agreement with European Guidelines, but this regulation is not yet operational. The amendment of the law of



11 December 1998 (23) left quite a lot of room for interpretation, because of the use of rather broad criteria. The draft of a Royal Decree for the implementation of the amendment to the law could clarify these issues. This draft is currently being finalised. The draft Royal Decree introduces a graduated regulation system for personal and medical data used for scientific research. It starts with anonymous data. If these data are not sufficient to achieve the aim, coded data can be used. Only if these two are still not sufficient to reach the aim uncoded personal data can be used. However, we believe that a separate legislation on cancer registration would be more appropriate (cf. Denmark).

## Conclusions

Cancer registration can be used to study the geographical distribution of cancer incidence in relation to environmental health factors. For this, the registry needs to contain detailed geographical information. The most appropriate level of registration for the study of a clusters seems to be the statistical code. If the health authorities consider this to be an important objective of a cancer registry, then a legislation on the registration of cancer is necessary; an amendment is not enough. A new legislation would lead to a more efficient organisation of cancer prevention programmes by means of a well-structured cancer registry compiled in an ethically, economically, and scientifically acceptable manner.

## Samenvatting

### *Doel van de studie*

Eén van de doelstellingen van kankerregistratie is het weergeven van de geografische spreiding van de nieuwe gevallen van kanker, wat, samen met andere informatie, kan gebruikt worden voor het bestuderen van milieugezondheidseffecten. Dit onderzoek wil nagaan of een kankerregister kan worden gebruikt voor het evalueren van milieugezondheidseffecten rond een puntbron en zo ja, aan welke vereisten het register hiervoor dient te voldoen. Verder wordt nagegaan in welke mate een kankerregister geografische gegevens dient te bevatten om milieugezondheidseffecten op een zinvolle manier te kunnen evalueren.

### *Belangrijkste resultaten*

Om bruikbaar te zijn voor onderzoek naar een kanker- of ziektecluster, dient een register aan een heel aantal voorwaarden te voldoen. Zo moet het bijvoorbeeld zo compleet

mogelijk zijn, accurate gegevens bevatten en dient men te beschikken over een procedure om dubbele registraties te voorkomen. Bovendien moeten de gegevens ook snel beschikbaar zijn. In Vlaanderen zou een netwerk van de bestaande kankerregistraties waarbij de gegevens worden gekoppeld een belangrijke meerwaarde kunnen betekenen voor het onderzoek naar milieugezondheidseffecten. Wanneer iemand een vermoeden van een ziektecluster rapporteert, gaat het vaak om een zeer lokale verhoging van de ziekte-incidentie. Daarom is het noodzakelijk dat het register meer gedetailleerde geografische informatie bevat. De meest geschikte registratie-eenheid hiervoor lijkt de statistische code te zijn. De inventarisatie van de incidentie gevallen van kanker op het niveau van de statistische eenheden is mogelijk, maar omwille van de huidige privacy-wetgeving zeer moeilijk en arbeidsintensief.

### *Belangrijkste besluiten*

Indien de overheid het weergeven van de geografische spreiding van de incidentie van kanker als zeer belangrijk onderschrijft, dan is een speciale wetgeving op de kankerregistratie noodzakelijk. Een aanpassing van de bestaande privacy-wetgeving is onvoldoende. Deze nieuwe wetgeving moet aanleiding geven tot een meer efficiënte organisatie van de preventie van kanker door middel van een goed opgebouwde kankerregistratie, waarbij de gegevens worden verzameld op een ethisch, economisch en wetenschappelijk aanvaardbare wijze.

## **Résumé**

### *But de l'étude*

Un des objectifs importants pour l'enregistrement de données sur le cancer est la présentation de la distribution géographique des nouveaux cas de cancer. Associée à d'autres informations, cette distribution permet d'étudier les effets de l'environnement sur la santé. Cette étude a pour objectif de vérifier si un enregistrement de données sur le cancer peut être employé pour évaluer les effets de l'environnement sur la santé autour d'une source de point et si tel est le cas, de déterminer les conditions qui doivent être remplies. Un autre objectif est d'examiner dans quelle mesure un enregistrement de données sur le cancer doit contenir de l'information géographique afin de servir à l'évaluation des effets de l'environnement sur la santé.

### *Principaux résultats*

Afin de permettre l'évaluation d'un agglomérat de cancer/maladie, un enregistrement doit répondre à un certain nombre de conditions. Par exemple, l'enregistrement doit être aussi complète que possible, apporter des données très précises et une procédure pour éviter un double enregistrement doit être disponible. De plus, les données doivent être disponibles rapidement. En Flandre, un réseau d'enregistrement des cancers existants, dans lequel les données seraient reliées, pourrait engendrer une importante plus-value pour l'étude des effets de l'environnement sur la santé. Quand une personne soupçonne l'existence d'un agglomérat de maladie, très fréquemment ceci ne concerne qu'une

surélévation très locale de l'incidence de la maladie. Pour cette raison, il est nécessaire que l'enregistrement contienne une information géographique plus spécifique. L'unité d'enregistrement la plus appropriée semble être le code statistique. L'inventaire de l'incidence du cancer au niveau des unités statistique est possible, mais très coûteux en main d'oeuvre à cause de la législation sur la vie privée.

### *Les conclusions principales*

Pour autant que les autorités considèrent importante la présentation de la distribution géographique des nouveaux cas de cancer pour l'étude des effets de l'environnement sur la santé une nouvelle législation sur l'enregistrement du cancer s'impose. Cette nouvelle législation doit permettre la création d'enregistrements de données sur le cancer contenant plus de détails géographiques, façon acceptable du point de vue économique, éthique et scientifique.

## References

1. DRIJVER M, MELSE JM. Ziekteclusters en milieuverontreiniging, een leidraad voor gezondheidsdiensten. T. Soc. Gezondheidsz. 1992; 70: 565-570.
2. ALEXANDER FE, WILLIAMS J, MC KINNEY P, RICKETTS T, CARTWRIGHT R. A specialist leukaemia/lymphoma registry in the UK. Part 2: Clustering of Hodgkin's disease. Br. J.Cancer, 1989; 60: 948-952.
3. ALEXANDER FE. Clustering and Hodgkin's disease. Br. J. Cancer, 1990; 62: 708-711.
4. ALEXANDER FE, MC KINNEY PA, CARTWRIGHT RA et al. Methods of mapping and identifying small clusters of disease with application to geographical epidemiology. Geog. Anal, 1991a; 23(2): 156-173.
5. BESAG J. Contribution to discussion. RSS meeting on leukaemia clustering. J R Stat Soc Series A, 1989; 152: 367.
6. CUZICK J, EDWARDS R. Test for spatial clustering in heterogeneous population. J R Stat. Soc Series B, 1990; 52, 73-104.
7. Committee on Medical Aspects of Radiation in the Environment. Investigation of the possible increased incidence of leukaemia in young persons near the Dounreay Nuclear Establishment. Caithness. Scotland HMSO, London, 1988.
8. URQUHART J, BLACK R, BUIST E. Exploring small area methods. In: Elliott P et al. Methodology of Enquiries into Disease Clustering. SASHU, London, 1989: 41-49.
9. KNOX EG. The detection of space-time interactions. Appl. Stat., 1964; 13, 25.
10. BARTON DE, DAVID FN, FIX E. Tests for Space-time Interaction and a Power Function. Proceedings of the 5<sup>th</sup> Berkeley symposium on mathematical statistics and probability. University of California Press, Berkeley, 1966.
11. EDERER F, MYERS MH, MANTEL N. A statistical problem in space and time: do leukaemia cases come in clusters? Biometrics, 1964; 20: 626.
12. ROTHMAN KJ. Modern Epidemiology. Boston/Toronto: Little, Brown and Company, 1986.
13. ELLIOTT P, WESTLAKE AJ, HILLS M, et al. The Small Area Health Statistics Unit: a national facility for investigating health around point sources of environmental pollution in the United Kingdom. J Epidemiol Commun H, 1992; 46: 345-349.

14. JOLLEY DJ, JARMAN B, ELLIOTT P. Socio-economic confounding. In: Elliott P, Cuzick J, English D et al. *Geographical and Environmental Epidemiology; Methods for Small-Area Analysis*. Oxford: Oxford University Press, 1997: 115-124.
15. SWERDLOW AJ. Cancer incidence data for adults. In: Elliott P, Cuzick J, English D, et al. *Geographical and Environmental Epidemiology. Methods for Small-Area Analysis*. Oxford: Oxford University Press, 1997: 51-62.
16. VAN EYCKEN E, HAUSTERMANS K. Kankerregistratie in Vlaanderen: wat, waarom en hoe? *Tijdschr Geneeskd* 2000, 56(4): 268-279.
17. TEPPLO L. Cancer registries in environmental cancer epidemiology. *Scand J Work Environ Health* 1998; 24: 1-2.
18. AUVINEN A, HAKAMA M, ARVELA H et al. Fallout from Chernobyl and incidence of childhood leukaemia in Finland, 1976-1992. *Brit Med J* 1994; 309: 151-154.
19. KRISTENSEN P, ANDERSEN A, IRGENS LM, et al. Cancer in offspring of parents engaged in agricultural activities in Norway: incidence and risk factors in the farm environment. *Int. J. Cancer* 1996; 65: 39-50.
20. WALLER LA, TURNBULL BW, GUSTAFSSON G, et al. Detection and assesment of clusters of disease: an application to nuclear power plant facilities and childhood leukaemia in Sweden. *Stat Med* 1995; 14: 3-16.
21. EHRENBERG L, EKMAN G, SVENSSON A. Epidemiological studies of geographic variations of cancer incidence in Sweden. *Acta Oncol* 1990; 29: 961-969.
22. Wet van 8 december 1992 tot bescherming van de persoonlijke levenssfeer ten opzichte van de verwerking van persoonsgegevens, B.S. 18 maart 1993, 5801.
23. Wet tot omzetting van de richtlijn 95/46/EG van 24 oktober 1995 van het Europees Parlement en de Raad betreffende de bescherming van natuurlijke personen in verband met de verwerking van persoonsgegevens en betreffende het vrij verkeer van die gegevens, B.S. 3 februari 1999, 3049-3065.