



Cancer in Norway 2010

2010

Cancer in Norway

Cancer incidence, mortality,
survival and prevalence in Norway

Special issue:
Clustering of cancer



Cancer in Norway 2010

Editor-in-chief: Inger Kristin Larsen

Analysis: Bjørge Sæther, Bjarte Aagnes

Layout and design: Gunther Zerener

Correspondence to: Inger Kristin Larsen - inger.kristin.larsen@kreftregisteret.no

Editorial team: Inger Kristin Larsen, Tom Kristian Grimsrud, Tom Børge Johannesen, Aage Johansen, Hilde Langseth, Siri Larønningen, Jan Ivar Martinsen, Christine Mellem, Bjørn Møller, Jan Franz Nygård, Bjørge Sæther, Svein Erling Tysvær, Gunther Zerener, Bjarte Aagnes, Giske Ursin

Recommended reference:

Cancer Registry of Norway. Cancer in Norway 2010 - Cancer incidence, mortality, survival and prevalence in Norway. Oslo: Cancer Registry of Norway, 2012.

Special issue: Clustering of cancer

Editor: Tom K Grimsrud

Writing group: Odd O Aalen, Tom K Grimsrud, Steinar Tretli

Layout and design: Gunther Zerener

Linguistic assistance: Barbara Mortensen

Correspondence to: Tom K Grimsrud – tom.k.grimsrud@kreftregisteret.no

Recommended reference: Cancer in Norway 2010. Special issue: Clustering of cancer. Grimsrud TK (Ed). Oslo: Cancer Registry of Norway, 2012

ISBN: 978-82-90343-3

ISSN: 0332-9631

General requests for cancer information, data or possible research collaborations are welcome, and should be sent to datautlevering@kreftregisteret.no

Special Issue: Clustering of cancer

Editor: Tom K Grimsrud

Writing group: Odd O Aalen, Tom K Grimsrud, Steinar Tretli

Layout and design: Gunther Zerener

Linguistic assistance: Barbara Mortensen

Correspondence to: Tom K Grimsrud – tom.k.grimsrud@kreftregisteret.no

Recommended reference: Cancer in Norway 2010. Special issue: Clustering of cancer. Grimsrud TK (Ed).
Oslo: Cancer Registry of Norway, 2012



Contents

– an historical approach –

Cancer clustering, a challenge for laymen and scientists 105

Tom K Grimsrud, Cancer Registry of Norway, Oslo, Norway

Abstract	105
1 Introduction	105
2 What is a cluster	105
3 Clusters pointing at causes	107
4 Identification of human carcinogens	109
5 Challenges and limitations in cluster inquiries	110
References	112

– a statistical approach –

Can we draw causal conclusions from unexpected clusters of disease?..... 115

Odd O Aalen, University of Oslo, Oslo, Norway

Steinar Tretli, Cancer Registry of Norway, Oslo, Norway

Abstract	115
1 Introduction	115
2 The concept of cause in disease clustering	116
3 The impact of randomness	117
4 Analysis of clusters to assess for randomness	119
5 The aspect of time	120
6 Variation in risk: Clustering as a signal	120
7 Power to detect anomalies.....	121
8 Final comments.....	122
References	123

– a practical approach –

Cluster inquiries, guidelines and lessons to learn 124

Tom K Grimsrud, Cancer Registry of Norway, Oslo, Norway

Steinar Tretli, Cancer Registry of Norway, Oslo, Norway

Tor Haldorsen, Cancer Registry of Norway, Oslo, Norway

Abstract	124
1 Introduction	124
2 Reported experience from cluster inquiries in the USA	125
3 Guidelines and information on cancer clusters	125
4 Perceived clusters addressed by the Cancer Registry of Norway	126
5 Cancer cluster resources	130
Abbreviations	131
References	132



Cancer clustering, a challenge for laymen and scientists

Tom K Grimsrud
Cancer Registry of Norway, Department for research

Abstract

The public concern following a clustering of cancer cases represents a challenge for health workers, media, and politicians. There is no easy way to sort random clustering from an effect of carcinogenic exposure. Historically, some cancer clusters have provided clues that later led to the identification of cancer causes. Most perceived cancer clusters, however, are assumed to be a result of random variation. Still, the worry and questions caused by a perceived cluster need to be handled carefully, with epidemiological skills, and public health experience. The paper describes why scientific expectations from present-day cluster investigations in general should be low.

1 Introduction

Most people correctly interpret a high frequency of disease or death as a possible sign of a health hazard. Everyone is familiar with the synchronous attacks of chickenpox, or the common fate shared by guests after food poisoning. This type of insight has been used as a guide for prevention of death and disease since ancient times. In modern western societies, monitoring of disease, accidents, and mortality—based on reporting of diseased individuals or aggregated data—is an important tool for protection of health and maintaining of security in the population.

Through the ages, the societal responsibility for correct handling of health threats has been transferred from religious authorities, to sovereigns assisted by astrologers, to health professionals, and ultimately to politicians. During recent decades, we have seen a growing public awareness of health issues, and the media involvement has increased. The necessary action against a health threat may be obvious in some situations, but often a fundamental uncertainty remains.

2 What is a cluster?

A confirmed disease cluster may be defined as “the occurrence of more than the expected number of people diagnosed with a certain disease within a specific group, a geographic area or a period of time”

(California DPH, 2012). The “expected” number is often derived from disease rates in the general population or another comparable group, usually adjusted for age distribution and sex.

Some illnesses, such as contagious diseases, may present a short time after exposure, which, combined with a usual large number of cases, easily will produce clusters. But even rare diseases with no common cause may be found to have a clustered distribution. In fact, it can be shown statistically that random and rare events—characteristic of many diseases—tend to occur irregularly, meaning that cases are not evenly distributed in time and space. The cases rather tend both to aggregate and to leave open spaces between them. A clustering of disease may therefore be observed even in the total absence of any common cause. This inherent link between randomness and irregular distribution is often perceived as counterintuitive.

Rothman defined a cluster as an aggregation of cases by season, year, time of day, sex, age, race, occupation, diet, or any environmental or genetic circumstance (Rothman, 1990). The approach alludes to a general discussion of contrast and variation, which is fundamental for all epidemiological research.

2.1 Cancer clusters

Chronic diseases may develop over time (years or decades) and they may have a complex set of causes. For some cancer forms, there is no known cause at all, and in a population, most cancers are explained only in part by known or suspected carcinogenic exposures. In the situation where a group of people are exposed to a common carcinogen, there may be different timing and degree of exposure, the disease may need additional exposures or the occurrence of random cell events to develop, and, ultimately, there may be differences in growth rate between tumours of the same type of cancer. Such a variation in timing and disease development will easily result in an apparent pattern of randomness, with cases occurring at different times and at different ages, despite the existence of a common cause.

Cancer may be seen as common and rare at the same time depending on the definition. All cancer diseases carry some common features such as uncontrolled growth of cells, and the tendencies to destroy adjacent tissue and spread to other organs or parts of the body. Still, a cancer can belong to one out of many rare types and subgroups according to its origin (organ, or part of the body), its appearance in the microscope, or the functional or genetic makeup. These quite different types of cancer usually have their own sets of risk factors.

Most cancers occur late in life (after 50), which leads to the fact that people over time normally face a marked increase in cancer incidence among their peers. About half of the cancer cases in a western population belong to the most frequent types, including cancers of the colon, rectum, lung, prostate, and female breast. The regular statistical reports from the Cancer Registry (Cancer in Norway, 2010) provide data for some 40 different types or subgroups of cancer. The cumulative risk for most cancer forms is less than 10 percent before age 75, and the occurrence of a single type of cancer therefore represents a quite rare event.

In total, cancer constitutes one of the most common causes of death. A potentially severe illness is more easily recalled by family, neighbours, and colleagues, and the dramatic aspect of losing a friend or family member may induce fear and increase the awareness of any sign of clustering. It may also stimulate the interest in avoidable exposures and prevention, which is often shared by media and the scientific audience. Norwegian doctors were engaged in these questions in the early 1900s, see Figure 1. Given the mechanisms for cancer development described

above, it is, however, fairly unlikely that a common cause should lead to synchronous outbreaks, even if the timing of the causal exposure is the same.

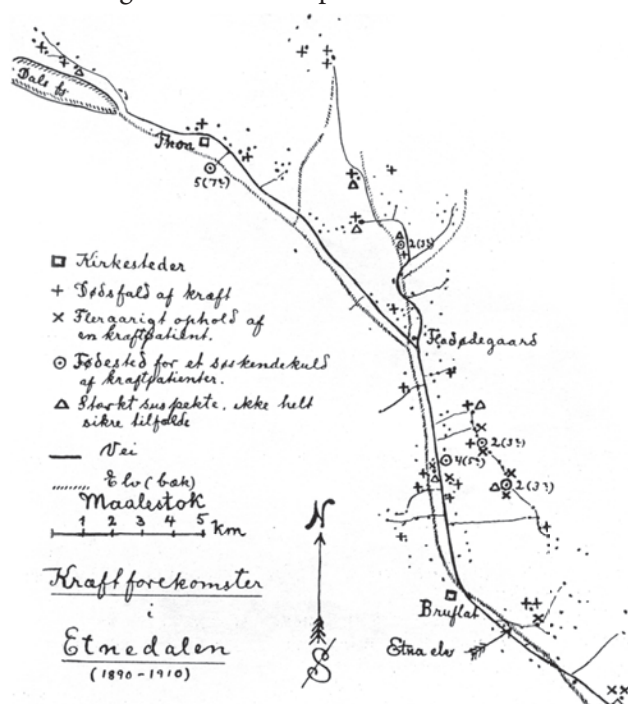


Figure 1. Mapping of cancer cases in the valley of Etnedalen 1890–1910. A hundred years ago, Norwegian doctors reported perceived cancer clusters in the Journal of the Norwegian Medical Association (Lunde, 1910)

The more remarkable it will appear when several cases of the same rare cancer are seen within a small geographical area or a short time interval, and even more so if the patients are young. A closer look into clusters of this kind would therefore be reasonable, even if we know that clustering commonly is a result of random events.

A cancer may spread (metastasise) to other parts of the body, such as lung, liver, brain, and skeleton, making it difficult to ascertain what type of cancer it is. Medical skills are therefore indispensable in the evaluation of a perceived cancer cluster, already from the descriptive stage.

2.2 Clustering of exposure

When health-care workers detect a cluster, often called a medical cluster, the apparent link to a common and localised *exposure* may be more striking than the aggregation of the disease itself. Typical examples are cases of food poisoning from a single party or restaurant, or traffic accidents occurring along a perilous stretch of a road.

Acute or short-term effects of a chemical or nuclear disaster will occur more closely in time than chronic

diseases. Health consequences in terms of cancer may develop over years or decades, and they can be difficult to detect, both among workers, rescue personnel, and local residents. In 1976, an unfortunate episode produced a cloud with very toxic chlorinated organic compounds (among them tetrachlorodioxins, TCDD) contaminating the local surroundings of a chemical factory in Seveso, Italy. Local inhabitants were highly exposed, but it was necessary with some 20 years of observation until an increased risk of cancer could be identified (Bertazzi & al, 2001).

Still, there are a number of historical examples of medical cancer clusters that have proven to be very useful for the understanding of cancer causes. In fact, a potential clustering of *exposure* may appear as the most interesting part of a cluster inquiry, both for the scientific community and for the public, as the underlying concern often includes idea that the disease incidents have a common origin.

2.3 Induction time and latency

The length of induction time (from exposure to onset of disease) or latency period (from start of disease to diagnosis) may vary according to the disease itself and the individual characteristics of the patient. Possibly, it may also vary according to the intensity and duration of the exposure. Some cancers may reach their peak occurrence within 2–15 years after exposure (leukaemia caused by radiation or benzene), while other cancers may develop up to 6 decades later (mesothelioma caused by asbestos). A long delay between exposure and disease may hamper the identification of an association.

3 Clusters pointing at causes

The frequency of illness has been commented on since Hippocrates (400 BC) (Merrill, 2010). Equally a long, high number of fatal or disabling diseases has generated fright, and increasingly so during times when there was no knowledge of what we would call a cause or a risk factor.

Through the last centuries, some observations of cancer clustering have prompted the search for an explanation. A few clusters eventually led to the identification of specific carcinogenic exposures, or a general description of conditions or activities associated with elevated risk.

155 DE MORBIS
hunc inter mammas & uterum admittere necessum est, experientia satis attestante, in mulierum mammis, ob uteri exorbitantias, generari perſæpe cancroſos tumores, quales in Monialibus magis, quam in cæteris Mulieribus, obſervantur, non ob menſtruorum defectum, ſed potius, ut reor, ob cælibem vitam; mihi enim ſæpius obſervare contigit, Veſtales Virgines bene coloratas, menſtruis purgationibus rite fluentibus, ſed ſalaci natura præditas, ex horrendis mammarum cancris miſere obiſſe; & quoniam in Italia quælibet Civitas complures hæbet Religioſos Virginum Cœtus, perraro fit, ut Monaſterium aliquod extet, quod tam diram peſtem intus non alat. Cur ergo propter uteri deliria plectuntur mammæ, non ſic, neque tam frequenter, aliæ partes? Certe ob conſenſum adhuc occultum, & Proſectorum indagini impervium, quem forſan dies aliqua aperiet, cum nondum ſit occupata veritas.

Figure 2. Bernardino Ramazzini (Ramazzini, 1703) wrote in his second edition of "Diseases of workers" about nuns (Monialibus, Vestales Virgines) suffering from dreadful breast cancers (horrendis mammarum cancris) assumed to be caused by a life in celibate (cælibem vitam), and commonly seen in convents in Italian cities. Courtesy of Internet Archive

The striking occurrence of breast cancer in Italian nunneries was noted by Ramazzini in the early 1700s (Ramazzini, 1703), see Figure 2. The identification of a "cause" in the sense of modern evidence-seeking medicine was, of course, far out of Ramazzini's reach, but in retrospect, he was probably facing a twin effect from bearing no children: The avoidance of childbirth eliminated a hazardous situation in most women's life (pregnancy and birth), and thereby led to a longer life and a higher risk of contracting any cancer. Later, bearing no child has been found to increase the risk of breast cancer, and the contrast between nuns and moms probably was more striking in earlier centuries when most women delivered a high number of children. Ramazzini's observation has proven to be in line with later findings (Fraumeni & al, 1969).

3.1 Important occupational cancer clusters

A number of other cancer causes have been detected as a result of occupational cancer clusters. Scrotal skin cancer in chimney sweeps appeared to be quite well understood by Pott when he published his observations in 1775 (Pott, 1775), although the suggested chemical origin of the disease was confirmed some 140 years later by the first successful experiment demonstrating an animal carcinogen by painting tar on the ears of rabbits (Yamagiwa & Ishikawa 1915).

In the nineteenth century, the deadly Schneeberg lung disease occurred with extreme frequency among miners in the small town Schneeberg in Saxony, Germany (Figure 3). It caused 50–75 percent of the deaths of local miners. Some victims were autopsied, and the disease was found to be a form of cancer in the lung (Härtig & Hesse, 1879). A general understanding of the true cause of this highly evident occupational and neighbourhood cancer cluster did not emerge until several decades later (Greenberg & Selikoff, 1993): The lung cancers were caused by extremely high levels of radon gas in the mines, emerging from uranium-containing ore.



Figure 3. The small town of Schneeberg in the Ore mountains (Erzgebirge), Saxony, Germany. Photo: André Karwath (Aka)

In 1895, three cases of cancer in the urinary bladder were reported in men younger than 50 years among 45 workers at a German plant producing fuchsine, also known as rosaniline hydrochloride, or magenta, an organic dye. The occurrence was taken to represent an occupational disease (Rehn, 1895). Additional epidemiological evidence came from UK in the 1950s, and later from Italy; and the manufacture of magenta is now considered carcinogenic for humans although the exact chemical to blame is not known (IARC, 1987; IARC, 2012).

A cluster of two cancers of the interior cavities of the nose was seen among nickel-refinery workers in South Wales in the 1920s (Doll, 1984). Soon after, more cases were reported in this quite obvious occupational cluster, but it took more than 60 years before nickel compounds were identified as the chemical cause of lung cancer and nasal cancer

(IARC, 1990). An investigation at a nickel refinery in Norway by the Cancer Registry was also prompted by a perceived cluster, and it offered important contributions to the evidence (Pedersen & al, 1973).

Women who painted watches and other equipment with a luminescent radium-containing material in the 1920s suffered a subsequent highly increased risk of bone sarcoma and cancer of the nasal sinuses (Brues & Kirsh, 1977). The carcinogenic effect of ionising radiation was already recognised, but the link between these cancer types and oral exposure had not been observed earlier.

3.2 Important non-occupational clusters

In the 1960s, a medical cluster of eight girls at the age of 15–22 were diagnosed with vaginal cancer, and a tiny case-control study revealed that the mothers had used hormone tablets (diethylstilbestrol) in order to protect against bleeding during pregnancy (Herbst & al, 1971).

In the early 1980s, a cluster of the rare cancer Kaposi's sarcoma appeared among young homosexual men with immune deficiency. Investigation of the cluster (Marmor & al, 1982) represented an important step towards identification of the AIDS syndrome, and Kaposi's sarcoma was later proven to be caused by human herpesvirus infection.

Investigation of clustering of cancer in families- preferably the same kind of cancer or combinations of certain types- has resulted in important knowledge of hereditary genetic conditions. During the 1990s, several genes responsible for inherited cancer risk were identified. The most common are the mismatch repair genes associated with Lynch syndrome (hereditary non-polyposis colorectal cancer, HNPCC) and the breast cancer and ovarian cancer genes BRCA1 and BRCA2 (Lynch & al, 2004). Familial clustering of cancer can probably be caused by other genetic risk factors than those identified to date, and by a combination of unknown genetic factors and environmental or lifestyle factors. Genetic mutation testing as a part of the health care system is available in Norway for persons with clustering of relevant cancers in their family.

The lifetime risk may be substantial in carriers of mutations in genes linked to hereditary risk of breast and ovarian cancer, some 10 times higher than that of the general population. On the population level, however, these conditions are rare, and estimated to explain less than 5–10% of the cases. Geographical clustering of mutation carriers has been reported by Norwegian geneticists (Møller & al, 2007).

We can conclude that historical examples from three centuries illustrate the role of cancer clustering for the advancement of human knowledge of cancer causes. For most cases, a long time elapsed between the initial suspicion and the recognition of a carcinogen. More precise characterisation of the causal associations has emerged with modern epidemiological methods developed after World War II, with increasing availability of high-quality animal experiments, or more recently, with biomolecular techniques.

3.3 What was special with the “useful” historical cancer clusters?

Three features were common for many of the informative historical cancer clusters. The absolute or relative risks were high, the exposures were rare and high, and some of the cancers (although not all) were rare. These qualities may all suggest an unusual hazard. The relative risk of a rare disease may reach higher levels than that of a more common disease, which can be helpful for the investigation. Among the Welsh nickel-refinery workers, the overall risk of the rare cancer of the nose and nasal sinuses was 200 times higher among exposed workers than expected from national rates (74 deaths observed against less than 1 expected) (Doll & al, 1990). Even higher relative risk estimates were reported in workers with extensive exposure.

A deviating age pattern among the cancer patients may also suggest an unusual hazard. Low age at diagnosis can be seen for hereditary cancers or diseases following exposures early in life. Normally, vaginal cancer affects women above 50, but rarely women in the age group typical for the diethylstilbestrol cases (Herbst & al, 1971).

Through the last decades, improved knowledge of cancer causes and an increasing interest in environmental issues may have escalated public awareness and lowered the threshold for reporting perceived clusters.

4 Identification of human carcinogens

Typically, a cluster investigation will be a single study within a small population. As such it has limited statistical power and restricted scientific impact. In recent years, most advances in the knowledge of human carcinogenesis have come from multiple large studies with good data on exposure, and results from laboratory analyses of chemical and biological samples. Some expert evaluations of potential human

carcinogens have relied heavily on mechanistic data and animal experiments. The WHO International Agency for Research on Cancer (IARC) has defined a set of criteria for classification of human carcinogens, found in the Preamble to its monographs (IARC, 2006) (see also Figure 4).

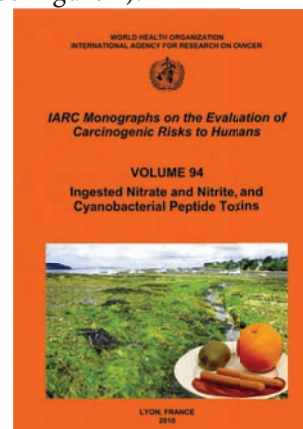


Figure 4. The International Agency for Research on Cancer (IARC) seeks to identify the causes of human cancer, and publishes authoritative monographs with evaluations conducted by international working groups (IARC, 2006).

In 1975, Doll reviewed known occupational cancer risks and concluded tentatively that only 4 out of more than 20 known carcinogenic exposures had been discovered in the course of initial evidence from laboratory experiments (Doll, 1975). In fact, observed clustering of cancer cases (or cancer deaths) in occupational groups continued to offer important clues and incentives for cancer research through the 1970s. For the role of asbestos in mesothelioma patients, the first described clustering included residents who had lived near open asbestos mines in childhood (Wagner & al, 1960).

Similarly, Neutra reviewed an updated list of recognised human carcinogens published by IARC in 1987. He found 26 out of 47 exposures or exposure situations to be discovered by medical or occupational clusters (Neutra & al, 1992). Only a single exposure was detected on the basis of a residential cluster alone (mesothelioma from erionite, a fibrous mineral used in house-building).

In line with these notions, the association between exposure to vinyl chloride monomer and angiosarcoma of the liver has been claimed to be detected as a cluster observation (Rothman, 1990). Strictly speaking, though, there were animal experiments that prompted the first evaluation of cancer risk among exposed workers (Doll, 1975; Maltoni & Lefemine, 1975). Still, it is fair to give the epidemiological results attention, as the relevance to humans always can be questioned for animal experiments.

4.1 Active search for cancer causes

A prerequisite for cluster evaluations, and a great help for aetiological cancer research, is the availability of background disease rates. An active search for cancer causes started in Norway in the last decades of the nineteenth century in the form of descriptive studies of the frequency of cancer deaths. The first attempt of establishing a national registration of cancer cases came in 1908 from a committee under the Norwegian Medical Association, led by the pathologist Fredrik Georg Gade. Cancer of the stomach was reported to be the dominating type of cancer in women and men combined, while breast cancer and gynaecological cancer were the second most frequent cancers in women (Gade, 1916).

A Norwegian national Cancer Registry was founded in 1952, with mandatory reporting of cancer from pathology laboratories and physicians. Valuable data on trends in incidence, on regional differences, and on survival have been produced during the subsequent six decades based on virtually complete data with increasing diagnostic ascertainment (Larsen & al, 2009).

Important contributions to the identification of cancer causes have been produced by Norwegian epidemiologists based on data from large population-based surveys and health examinations, data on occupational exposures from industrial plants, workers' unions, official statistics, and census data, all with subsequent linkage to the Cancer Registry data base.

5 Challenges and limitations in cluster inquiries

The reactive character of a cluster inquiry is not a good starting point for a scientific evaluation. Good epidemiological science should ideally start with a hypothesis—such as a suggested carcinogen or a risk factor—and proceed with the choice of the best population to test the hypothesis. Subsequently, one should choose a reliable way to identify cases in the study group and in the reference population, find means to avoid selection and misclassification, and collect data to obtain control for potential confounders that may disturb the statistical analysis.

A cluster inquiry, on the other hand, is often driven by a local request for an explanation of a perceived local excess of cancer. The expectations may create an unfavourable situation in a number of respects.

5.1 Study size

The problem with undersized studies has already been commented on. Typically, a cancer cluster is based on low numbers, often fewer than 10 cases, and the study population would not be selected after an independent evaluation, but rather suggested by the occurrence of the cases in time and space. Additionally, it is often difficult to decide the exact outline of the source population—or study population—considered to be relevant or “at risk”.

If the area or period defined by the perceived cluster is small a clustering of cases produced by random will give a falsely inflated (artificially high, not representative) relative risk. On the other hand, a comprehensive population or longer period of observation may dilute the risk estimate, and these issues may be highly controversial if a strong local cause is suspected by the public. Additionally, some of the common statistical tools meant for testing of hypotheses are not appropriate and remain less useful when there is an *a priori* knowledge or suspicion of a cluster. One should therefore bear in mind that clustering of cases also can be an indication of a generally increased risk in a larger area or a larger population group, and that it can be appropriate to widen the scope.

Epidemiological studies rely heavily on comparisons of groups of people. In most situations it is impossible to determine the cause of disease for a single individual or even a small group, unless there is a highly specific exposure with a strong effect, such as the biological agent for an infectious disease. The number of cases included in a study is the main factor restricting statistical power, and important comparisons between subgroups according to different levels of exposure may be impossible when too few cases are included. A small study of weak carcinogens is bound to be uninformative.

Interestingly, Norwegian geneticists have shown that restricting genetic testing to families with observed clustering of breast and ovarian cancer limits the detection rate to less than 50% of all BRCA mutation carriers in the population (Møller & al, 2007). Their findings illustrate the fact that a quite high relative risk in small groups (families) may materialise as clusters, but the identification and distribution of the clusters are an unreliable measure of the size of the problem at the population level. The parallel to other cluster inquiries is evident: Although a small cluster may need attention, it can be important and rewarding to allocate resources to well-designed large studies.

5.2 Diagnostic considerations

The knowledge of cancer causes is not equally distributed between the different types of cancer. For many of the common cancers—or groups of cancer—such as female breast, large bowel, prostate, and lympho-hematopoietic cancers, the proportion explained by known external or environmental exposures is low. Prostate cancer and large bowel cancer suffer from a striking paucity of recognised causes. A well-known exception is lung cancer, for which smoking tends to overshadow other potentially important causes such as occupational exposures, indoor radon exposure, and air pollution.

Some carcinogens may increase the risk of two or more different cancer forms. This is seen for tobacco (smoking and smokeless tobacco, snus), ionising radiation, arsenic, asbestos, alcohol, dioxins, and nickel compounds. Still, the size of the risk may vary between the cancer types. When facing a perceived cancer cluster, it is often useful to discuss the cancer pattern along with known explanatory factors.

Historically, the clusters that proved to be scientifically rewarding, were often picked up by health professionals with knowledge of background rates, either general practitioners, hospital doctors, researchers, or public health officers. We should also be aware the historical examples of serious delays in the identification of carcinogens and unreasonable rejection of new hypotheses. This has been seen both from the medical establishment (tobacco smoking and lung cancer) (Wynder, 1986) and from industry with strong economical interests (Michaels, 2008).

In advanced stages, a cancer disease may change its appearance. Tumour cells may spread to other organs or parts of the body, leaving it difficult to decide in what organ the cancer originally developed. In some cluster situations, local people may have collected individual information on disease and death, while the determination of exact diagnoses and identity remains a demanding task with erroneous and incomplete data. Additionally, ascertainment of cancer diseases by means of registry data may require permission from ethical boards for research, which may be reluctant to licence a public health activity with only questionable scientific value.

5.3 Scarce exposure information

Sometimes a cluster inquiry comes from an area where a potential health hazard from pollution has been debated for years. A cluster may also emerge

where no cause has been suspected. In the latter situation, a long list of proposed causes may exist. It is often useful to discuss exposure and disease at a general level, but an in-depth evaluation requires good exposure data of candidate causes and potential confounders. The quality of the exposure information is often a limiting factor in an epidemiological study, both for the validity and for the possibility to explore dose-response relationships.

Some of the most frequent non-occupational environmental carcinogens are ultra-violet radiation, some virus infections, in-door radon from the soil, and ambient air pollution. Potential confounding from exposures linked to lifestyle may be difficult to control, the most important ones being use of tobacco, alcohol consumption, excessive sunbathing, sexual activity, and handling of carcinogens associated with house-building materials. Some lifestyle habits are known to be unevenly distributed, sometimes according to socio-economic status, and they may vary with time, fashion, social norms, and with place of residence.

5.4 Bias

If a potential hazard or health problem has received much attention it can be difficult to collect unbiased information from the population. The situation may be particularly challenging if no independent data are available neither for the health outcome nor for the relevant exposures.

Sometimes outcome and exposure data have been collected from the same individuals by interview or questionnaire. This approach gives ample room for selection bias in participation, for recall bias, reporting bias, and attribution that can mimic or mask any true association. Claims and expectations of compensation might further compromise the validity of survey data, and can easily introduce distrust in the process. News media have an interest of their own to keep up the drama, discussion, and conflict of interest, and by these mechanisms, a suggested causal relationship in a perceived cluster may remain almost inaccessible for research—unless independent sources exist for exposure and health outcome.

Other expectations—in terms of a specific wanted explanation, a request for clean-up of local sources of pollution, or even improved house prices—can also make it hard for health professionals to establish a climate with enough trust from local inhabitants, enough empathy towards patients and relatives, and a satisfactory distance for an independent evaluation. When a question of clustering is linked to a political

agenda transparency and good communication skills are mandatory. The process may be time-consuming, but time and attention are important remedies that may help to save resources in the long run.

5.5 Science or social mission?

A successful handling of a perceived cluster would—as seen from the health care system—communicate a balanced picture of potential health threats, meaning that unreasonable fright is reduced, and focus is directed towards the most important determinants for health and disease. Skills and experience, available data sources, appropriate timing, and an empathetic approach can improve the process. Hurt feelings and

distrust in the population, or even mindless presentation by a news agency can increase toll and stress. It may be beneficial for all parties if a cluster is reacted upon properly from the beginning, in the sense that the basic aspects and questions are addressed with an evidence- and knowledge-based attitude, calling for a combination of scientific and public health approach. Lastly, on the national scale, one should remember that out of the 28 000 new cases of cancer registered in Norway every year, only a small proportion is perceived to emerge as clusters. Still, there are ample contrasts within the population, as well as changing time trends, that could be worthy of a closer description or an aetiological pursuit.

References

- Bertazzi PA, Consonni D, Bachetti S, Rubagotti M, Baccarelli A, Zocchetti C, Pesatori AC. Health effects of dioxin exposure: a 20-year mortality study. *Am J Epidemiol* 2001; 153(11):1031–1044. <http://dx.doi.org/10.1093/aje/153.11.1031>
- Brues AM, Kirsh IE. The fate of individuals containing radium. *Trans Am Clin Climatol Assoc* 1977; 88:211–218. <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2441427/?tool=pubmed>
- California DPH, California Department of public health. Cancer clusters. http://www.ehib.org/topic.jsp?topic_key=72
- Cancer in Norway, 2010. Oslo: Cancer Registry of Norway, 2010:p1–100.
- Doll R. Nickel exposure: a human health hazard. *IARC Sci Publ* 1984; (53):3–21.
- Doll R. Part III: 7th Walter Hubert lecture. Pott and the prospects for prevention. *Br J Cancer* 1975; 32(2):263–72. <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2024827/?tool=pubmed>
- Doll R, Andersen A, Cooper WC, Cosmatos I, Cragle DL, Easton D, Enterline P, Goldberg M, Metcalfe L, Norseth T, Peto J, Rigaut J-P, Roberts R, Seilkop SK, Shannon H, Speizer F, Sunderman FW Jr, Thornhill P, Warner JS, Weglo J, Wright M. Report of the International Committee on Nickel Carcinogenesis in Man. *Scand J Work Environ Health* 1990; 16(1 Spec No):1–82. http://www.sjweh.fi/show_abstract.php?abstract_id=1813
- Fraumeni JF Jr, Lloyd JW, Smith EM, Wagoner JK. Cancer mortality among nuns: role of marital status in etiology of neoplastic disease in women. *J Natl Cancer Inst* 1969; 42(3):455–468. <http://dx.doi.org/10.1093/jnci/42.3.455>
- Gade FG. Undersøkelser over kræftsygdommene i Norge, paa grundlag av den officielle mortalitetsstatistik 1902–1911 samt det av den norske komité for kræftforskning samlede materiale 1908–1912 [Investigations on cancer diseases in Norway, based on official mortality statistics 1902–1911 and material collected by the Norwegian Committee for cancer research 1908–1911]. Kristiania (Oslo): Dybwad, 1916:p1–102. [In Norwegian] <http://archive.org/stream/skrifterutgitavv161chri#page/n5/mode/2up>
- Greenberg M, Selikoff IJ. Lung cancer in the Schneeberg mines: a reappraisal of the data reported by Harting and Hesse in 1879. *Ann Occup Hyg* 1993; 37(1):5–14. <http://dx.doi.org/10.1093/annhyg/37.1.5>

Härting FH, Hässe W. Der Lungenkrebs, die Bergkrankheit in den Schneeberger Gruben [Lung cancer, the disease of miners in the Schneeberg mines]. *Vjschr gericht Med öff Sani Wes* 1879; 30:296–309; 31:102–129; 31:313–327. [In German]

Vol 30, p296: <http://archive.org/stream/vierteljahrssch09unkngoog#page/n305/mode/1up>

Vol 31, p102: <http://archive.org/stream/vierteljahrssch10prusgoog#page/n113/mode/1up>

Vol 31, p313: <http://archive.org/stream/vierteljahrssch10prusgoog#page/n324/mode/1up>

Herbst AL, Ulfelder H, Poskanzer DC. Adenocarcinoma of the vagina. Association of maternal stilbestrol therapy with tumor appearance in young women. *N Engl J Med* 1971; 284(15):878–881.

<http://dx.doi.org/10.1056/NEJM197104222841604>

IARC, International Agency for Research on Cancer. IARC Monographs on the Evaluation of Carcinogenic Risks to Humans, PREAMBLE. Lyon (France): WHO/IARC, 2006.

<http://monographs.iarc.fr/ENG/Preamble/CurrentPreamble.pdf>

IARC, International Agency for Research on Cancer. Overall Evaluations of Carcinogenicity: An updating of IARC monographs volumes 1 to 42. (Magenta and manufacture of magenta.) Lyon: IARC; 1987. IARC monographs on the evaluation of the carcinogenic risks to humans, suppl 7:238–239.

<http://monographs.iarc.fr/ENG/Monographs/suppl7/Suppl7-99.pdf>

IARC, International Agency for Research on Cancer. A Review of Human Carcinogens. Part F: Chemical Agents and Related Occupations. (Magenta and magenta production.) Lyon: IARC; 2012. IARC monographs on the evaluation of carcinogenic risks to humans, vol 100F:105–110.

<http://monographs.iarc.fr/ENG/Monographs/vol100F/mono100F-13.pdf>

IARC, International Agency for Research on Cancer. Chromium, nickel and welding. Lyon: IARC; 1990. IARC monographs on the evaluation of carcinogenic risks to humans, vol 49:257–445.

<http://monographs.iarc.fr/ENG/Monographs/vol49/index.php>

Larsen IK, Småstuen M, Johannesen TB, Langmark F, Parkin DM, Bray F, Møller B. Data quality at the Cancer Registry of Norway: an overview of comparability, completeness, validity and timeliness. *Eur J Cancer* 2009; 45(7):1218–1231. <http://dx.doi.org/10.1016/j.ejca.2008.10.037>

Lunde N. Litt om kræftens aarsaksforhold. (A little about causes of cancer.) *Tidsskr Nor Laegeforen* 1910; 30(10):426–430. [In Norwegian]

Lynch HT, Shaw TG, Lynch JF. Inherited predisposition to cancer: a historical overview. *Am J Med Genet C Semin Med Genet* 2004; 129(1):5–22. <http://dx.doi.org/10.1002/ajmg.c.30026>

Maltoni C, Lefemine G. Carcinogenicity bioassays of vinyl chloride: current results. *Ann N Y Acad Sci* 1975; 246:195–218. <http://dx.doi.org/10.1111/j.1749-6632.1975.tb51094.x>

Marmor M, Friedman-Kien AE, Laubenstein L, Byrum RD, William DC, D'onofrio S, Dubin N. Risk factors for Kaposi's sarcoma in homosexual men. *Lancet* 1982; 1(8281):1083–1087.

Merrill RM. Introduction to Epidemiology. Fifth edition. Sudbury (MA, USA): Jones & Bartlett Publishers, 2010:p24. ISBN 0-7637-6622-4.

http://books.google.no/books?id=RMDbh6gw1_UC&pg=PA24&dq=&hl=no#v=onepage&q&f=false

Michaels D. Doubt is their product: how industry's assault on science threatens your health. New York: Oxford University Press, 2008:p1–372.

Møller P, Hagen AI, Apold J, Maehle L, Clark N, Fiane B, Løvslett K, Hovig E, Vabø A. Genetic epidemiology of BRCA mutations – family history detects less than 50% of the mutation carriers. *Eur J Cancer* 2007; 43(11):1713–1717.

- Neutra R, Swan S, Mack T. Clusters galore: insights about environmental clusters from probability theory. *Sci Total Environ* 1992; 127(1-2):187–200.
- Pedersen E, Høgetveit AC, Andersen A. Cancer of respiratory organs among workers at a nickel refinery in Norway. *Int J Cancer* 1973; 12(1):32–41. <http://dx.doi.org/10.1002/ijc.2910120104>
- Pesatori AC, Consonni D, Rubagotti M, Grillo P, Bertazzi PA. Cancer incidence in the population exposed to dioxin after the “Seveso accident”: twenty years of follow-up. *Environ Health* 2009; 8:39.
- Pott P. Chirurgical observations. Vol 3. London: L Hawes, W Clark, and R Collins, 1775:p177-83.
- Ramazzini B. De morbis atrificum [Diseases of workers]. Second edition. Padova (Italy): Archi-Lycæo, 1703:p156. [In Latin] <http://archive.org/stream/bernamazzinide00porzgoog#page/n159/mode/1up>
- Ray M. Merrill. Introduction to Epidemiology. Fifth edition. Sudbury (MA, USA): Jones & Bartlett Publishers, 2010:p24. ISBN 0-7637-6622-4.
http://books.google.no/books?id=RMDBh6gw1_UC&pg=PA24&dq=&hl=no#v=onepage&q&f=false
- Rehn L. Blasengeschwülste bei Fuchsin-Arbeitern [Cancer of the urinary bladder in fuchsin workers]. *Arch Klin Chir* 1895; 50:588–600 [In German]
- Rothman KJ. A sobering start for the cluster busters’ conference. *Am J Epidemiol* 1990; 132(1 Suppl):S6–S13.
- Wagner JC, Sleggs CA, Marchand P. Diffuse pleural mesothelioma and asbestos exposure in the North Western Cape Province. *Br J Ind Med* 1960; 17:260-271.
<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1038078/pdf/brjindmed00208-0014.pdf>
- Wynder EL. Preventing disease. *JAMA* 1986; 255(8):1017–1017.
- Yamagiwa K, Ichikawa K. Experimentelle Studie über die Pathogenese der Epithelialgeschwülste [Experimental study of the pathogenesis of epithelial tumours]. *Mitt Med Fak Tokyo* 1915; 15:295–344. [In German]



Can we draw causal conclusions from unexpected clusters of disease?

Odd O Aalen^{a,b}, Steinar Tretli^b

^aDepartment of Biostatistics, Institute of Basic Medical Sciences, University of Oslo, Norway

^bCancer Registry of Norway, Department for research

Abstract

Clusters of disease are often brought to the attention of health authorities. Some of these cases appear in the media, and create quite a stir. How do we interpret disease clusters in an epidemiological and statistical setting? Many disease clusters are due to random variation and we demonstrate the ability of randomness to produce clustering. When evaluating a cluster, Bonferroni type corrections can be made, but the degree of correction is often arbitrary and can only give a rough idea about whether the cluster in question could emerge by chance. A specific and localized cause could lead to clustering if the development of the disease has a short duration. Whether one should expect clustering in time depends on the length of the latency period of the disease, but also on variation in the latency period between individuals. For cancer, latency time is typically long, and usually one would not expect clustering of cases in time even in the presence of a common cause. When evaluating clusters there is a strong tendency to look for causal agents in the local environment. However, it may be more likely that the cluster points to a smaller increase of risk in a larger part of the population.

1 Introduction

Clusters of disease cases are a common phenomenon. These clusters may be found in neighbourhoods, at the work place or in the local community. It is often tempting to interpret a disease cluster as caused by an increased risk for a sub-population, due to specific chemicals, radiation or other exposures, affecting this sub-population. Still, in many of the supposed cluster situations it is more or less impossible to find a common cause. A critical review of clusters was given by Rothman (1990). A recent assessment was given in an editorial in *The Lancet Oncology* (2009), where they point out the fact that most cancer clusters will be random occurrences, but that one still has to be on the lookout for the rare cases where there could

be some substance behind the findings. Also, there is a responsibility to the public to make a thorough evaluation when an apparent cluster causes great attention. We would like to point out that there is a further aspect, namely that a cluster, although a random occurrence, may still signal a moderate underlying increase in incidence.

The issue of cancer clusters often occurs in the news media and popular literature, and a somewhat sophisticated discussion was given in *The New Yorker* (Gawande, 1999). A Norwegian disease cluster that has been given much attention over several years was the occurrence of congenital anomalies in the offspring of personnel serving aboard a missile torpedo boat (Magerøy & al, 2006).

In the present paper we consider clusters that are unexpected. Clusters that arise around a focus point where there exists a prior suspicion of increased risk are not the subject of this study. An example of the latter type might be clustering of leukaemia cases close to an atomic processing plant, like Sellafield (Draper, 1993).

An unexpected cluster will typically be informally detected by individuals in a local environment and then occasionally be reported to the health authorities, for example a cancer registry. These clusters are usually not well documented by the observer, and different interests, motives and limited knowledge about the disease might be the reason for the claim.

Most events within such clusters would not only be geographically close, but also close in time. In fact, the closeness in both place and time is often a prerequisite for the cluster to become an object of attention because the observer needs a reasonable overview of the population where the cluster is observed.

An evaluation of a cluster situation will therefore include several steps: 1. Are the observed number of cases more than expected in a comparative reference population? 2. What could be the reason for a changed disease incidence rate? 3. Is the suggested reason plausible? 4. Do other observations exist that support the explanation of the cluster occurrence?

For most disease clusters one would imagine a chemical, physical or biological cause, if there is any common cause present at all. Speculations often comprise effects of human activity such as radiation. It should be noted, in passing, that also clusters of a different nature exist. Internationally there are several known cases where clusters of deaths in nursery homes or hospitals have led to suspicions of criminal acts, and indeed nurses, in particular, have been charged with murder, see the Nature paper by Buchanan (2007). The considerations in these cases have similarities with evaluations of ordinary disease clusters, for example cancer clusters, but they also clearly have their own aspects. In particular, one would need far stronger evidence to conclude that there is a real cluster.

In the present paper we shall discuss the concept of cause in clustering and the possibilities and power to detect anomalies.

2 The concept of cause in disease clustering

The fundamental issue in assessing disease clusters is to judge whether there could be some common cause behind the cases that constitute the cluster. But it is not immediately obvious what is meant by cause in this context and we shall discuss certain aspects of this.

The situation is closely connected to epidemiological association studies where we attempt to decide whether an observed association between exposure and disease is causal. A set of criteria is usually considered (Hill, 1965) and a couple of them might be informative in the discussion of causality in the cluster context.

2.1 Hills criteria

One of the criteria is *biological plausibility*. This might be more important in the causality discussion of clusters than in epidemiological association studies. This is because the typical epidemiological studies are carried out by researchers who have considerable knowledge about the disease in question, and the biological plausibility is a part of the hypotheses that form the basis of the study. Disease clusters, which are often claimed by people without professional medical knowledge, may include different sub-diagnoses which are connected to different known risk patterns. Cancer is an example: The outcome variable(s) in an epidemiological study is usually one or a few specific types of cancer while in many reported unexpected clusters all types of cancer could be included. Therefore, a discussion of the plausibility of a cluster will also include whether the cluster situation is in conflict with what is known about risk factors for the involved types of cancer and what is known about the cancer disease process -especially the length of the process and the variation between individuals (this is part of the *coherence* criterion). Biological plausibility is also related to the concept of mechanistic understanding. Do we have any sense of the mechanisms that might lead to a cluster?

Another of Hill's criteria is *temporality*. In a discussion of causality, it is obvious that the cause shall precede the effect in time. The same necessary temporality exists for clusters, but the time frame is not always clear. Development of cancer, for instance, is a long process and it may not be clear when in the disease process the exposure has an effect. The authors' experience is that the general public tend to look for exposure

taking place in a rather short time before the cluster is claimed, and hence the suggested cause may not be realistic in the sense of temporality.

2.2 Counterfactual causality

The modern literature of causal inference is largely based on a counterfactual causality concept. This means that there is one factor that can be changed, keeping everything else equal, and such that one sees a clear difference in effect. When disease clusters are observed there is often a strong suspicion of a specific culprit. An example is given by Gavin and Catney (2006) where local concern over a cancer cluster in Northern Ireland lead to the unauthorized felling of a telecommunications mast. The authors show that in actual fact the incidence and mortality of cancer were within or lower than expected and so the suspicion harboured by members of the community was not justified.

Cancer clusters only rarely lead to the discovery of a specific counterfactual cause. One example is the cluster of 25 cases of mesothelioma in the Turkish village Karain, which turned out to be due to a mineral (erionite) and possibly asbestos as well (Honjo & al, 1982).

Even if there is a real increase in incidence, there may not be any clear counterfactual cause of this. The fact is that for a number of reasons the incidence of disease such as cancer varies between geographical areas, communities, professions and other groups. We shall show below that this could produce clusters, but there would be no one specific cause of this, but rather a complex web of causal elements combined with randomness.

There is a distinction in the causal literature between necessary and sufficient causes (Rothman & al, 2008). As regards clustering, does one imagine that a specific necessary cause exists in the local community, that is, a condition or an event, without which none of the cases would have arisen? This could for instance be a strong source of pollution, such as radiation or asbestos. Or does one mean a factor that increases the risk without being a necessary cause, such that the disease cases could have arisen anyway but that the factor in question increases the risk of this happening? Such a factor could be a part of a multifactorial set of causal influences where it might be impossible to say which factor is the major cause in a given case.

This is also related to the issue of insurance or compensation which would sometimes be an aspect

of the investigation. In order to give compensation, it is not a necessary requirement that one can prove that one specific cause is the guilty one. Rather one would make statistical requirements, for instance that the risk of disease is at least twice as large as it would have been without the putative cause, see for example Palmer & al (2007).

3 The impact of randomness

3.1 Random clustering

It is important to understand the power of randomness to induce clustering between events that are independent of one another, see also Aalen (1986). Independent events do not occur in a regular manner; rather the natural irregularity of randomness produces, so to speak, accidental clusters. This can be demonstrated by mathematical calculations or by simulation.

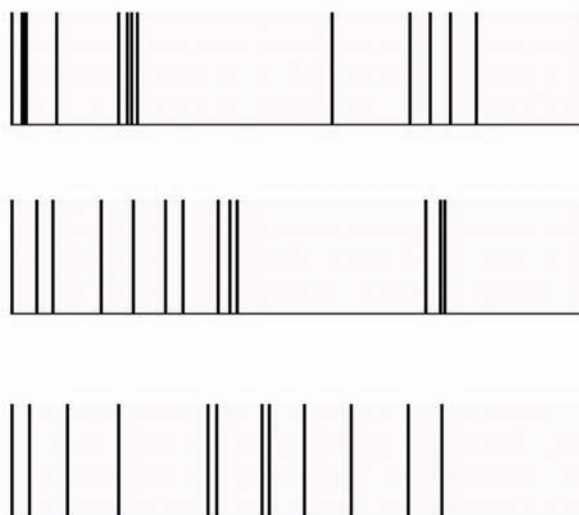


Figure 1. Simulation of a Poisson process with a constant rate. The horizontal line is the time axis, divided into three parts. The vertical line pieces indicate occurrence of events.

The Poisson process is the mathematical description of events occurring randomly and completely independent of one another. A simulation of a Poisson process with a fixed rate is demonstrated in Figure 1. The bottom line indicates time, divided into three panels, and the bars indicate the occurrence of events. One clearly sees the clustering. Indeed, the second bar in the top panel consists of two or three events that are almost simultaneous, and also in the rest of the figure there is considerable clustering and long and empty intervals in between. Figures 2 to 4 give similar demonstrations with a Poisson process over the plane. Again one sees the clustering and

the empty spaces in between. In Figure 4 the plane is divided in 25 grids with an expected rate of two events per grid. The observed occurrences in the grids vary from 0 to 6 occurrences. These simulations demonstrate that a considerable degree of clustering would be a natural phenomenon due to pure randomness, and that an understanding of this issue is a prerequisite for analyzing disease clusters.

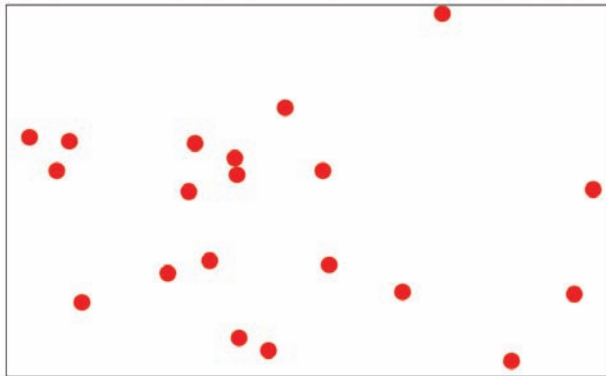


Figure 2. Simulation of a Poisson process over a plane with 20 points

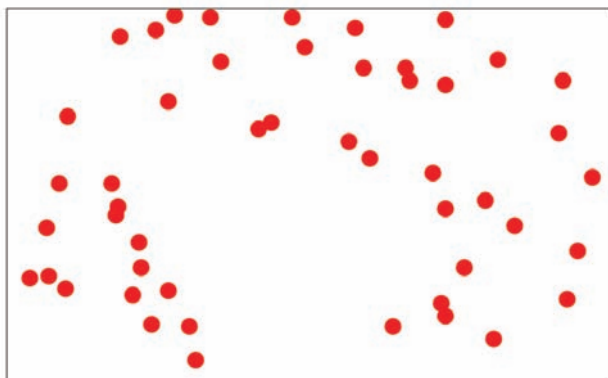


Figure 3. Simulation of a Poisson process over a plane with 50 points

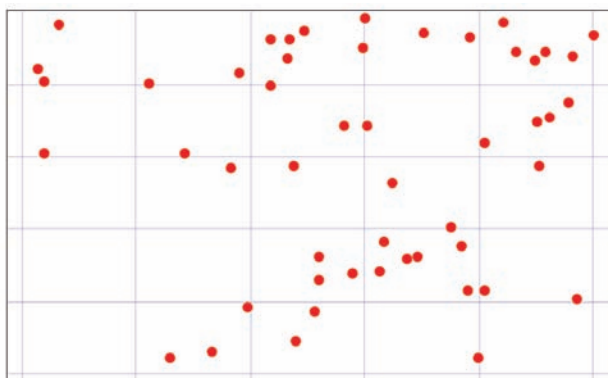


Figure 4. Simulation of a Poisson process over a plane with a grid

3.2 On the maximum of a set of observations

The observations that will cause special attention and give a basis for further investigations will be those that deviate strongly from what is expected. Usually the maximal number will be of greatest interest. We imagine that we observe occurrences of a rare condition over numerous small areas, and will consider the maximum number of occurrences in any single area. As a matter of simplicity we shall imagine that all areas are of the same size. The expected number, or rate, for each area is denoted λ . If the numbers of occurrences in different areas are independent, then the occurrences will form a Poisson process.

For those interested in the mathematics, the relevant formulas are as follows (standard probability theory): The point probability of a Poisson distribution (that is the probability of observing exactly x occurrences within a specific area) is given by $\lambda^x e^{-\lambda} / x!$ with cumulative distribution function

$$\sum_{y=0}^x \frac{\lambda^y}{y!} e^{-\lambda} = \frac{\Gamma(x+1, \lambda)}{x!}$$

where $\Gamma(x+1, \lambda)$ is the incomplete gamma function.

Consider n independent areas, each with a Poisson distributed number of occurrences with rate λ . Let U be the observed maximum number of occurrences (cases) in a single area. The cumulative distribution function of U follows from the Poisson distribution formula, and is given by

$$F(u, \lambda, n) = P(U \leq u) = \left(\sum_{y=0}^u \frac{\lambda^y}{y!} e^{-\lambda} \right)^n = \left(\frac{\Gamma(u+1, \lambda)}{u!} \right)^n$$

where u is a fixed value of the random variable U , λ is the rate, and n is the number of disjoint areas.

The point probability is

$$P(U = u) = F(u, \lambda, n) - F(u-1, \lambda, n)$$

and the expectation is

$$E(U) = \sum_{u=0}^{\infty} (1 - F(u, \lambda, n)).$$

Figure 5 presents the expectation of the maximum number of occurrences (cases), U , for various values of n and λ . The figure shows that the expected maximum is large compared to the expected number per area. For instance, the latter may equal 1, and still the expected maximum is almost 6 for 1000 areas. So, in a mathematical way, this shows that the maximum number of occurrences (cases), which is likely to get attention, may by pure chance be much higher than the average number.

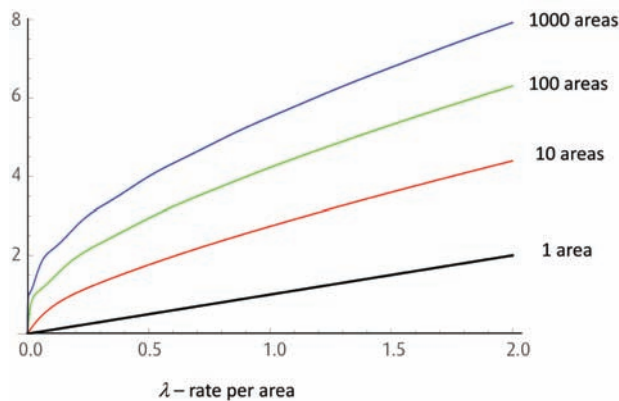


Figure 5. The expectation of the maximum number of occurrences (cases) over n disjoint areas, with n equal to 10, 100 and 1000.

4 Analysis of clusters to assess for randomness

4.1 Cluster of stillbirths

This example is inspired by a real occurrence, but modified somewhat. Although it is not about cancer, it illustrates important principles. During a five year period ten employees in a kindergarten became pregnant. Three of them had a stillborn child. This was a striking occurrence in a small group of colleagues and was reported to the authorities. A question was whether this could be due to something in the environment. There was no particular theory for how this could come about and statistical advice was sought in order to evaluate the likelihood that it could be a random occurrence. We shall do some calculations for this example.

We assume the probability of stillbirth to be 1% (which was a realistic value in the 1980s when this cluster was observed). We can then compute the probability that 3 out of the 10 pregnancies give a stillborn child:

$$\binom{10}{3} \times 0.01^3 \times 0.99^7 = 0.00011 = \underline{0.011\%}$$

4 out of 10 cases are even much less likely and when added to the above does not change the result and it turns out that the p-value equals 0.011 % for the occurrence of 3 or more stillborn.

Here we have computed a p-value for a hypothesis that was suggested by the data. One way to handle such post hoc hypotheses is to use a Bonferroni type correction. The idea is that such a cluster will be brought to attention if it is observed in one out of a large number of kindergartens or other small offices or working places of about same size and where the employers knew each other well enough to have the information that a college has received a diagnosis. The big question is how many units (kindergartens and possibly other work places) that one should reasonably consider within a time period and a geographic defined area. We rather arbitrarily choose 1000. The probability that at least 3 stillbirths occur one or more times among 10 employees would then be

$$1 - (1 - 0.00011)^{1000} = 0.10 = 10\%$$

However, it could easily be argued that 1000 could be increased to 5000 and then the probability would be about 42%. A look into the post hoc situation often changes the view that the unexpected cluster is very striking. It is, however, not so easy to communicate this view to the general population, and it typically remains uncertain how many "similar units" one should correct for.

4.2 Cancer cluster

A cancer cluster was observed among women working in a Norwegian library. During a period from 1980 to 1997 five cases of breast cancer were diagnosed in female employees. The total number of person years under risk was 818, and from Norwegian general population cancer data, considering the ages of the employees, the expected number of cases among employees at the library was 0.73. Hence, the observed to expected ratio was $5/0.73 = 6.85$. The p-value, that is the probability of observing 5 or more cases when the expected number is 0.73, may be calculated from a Poisson distribution yielding $0.00094 = 0.094\%$. Considering 1000 similar units gives Bonferroni adjusted p-value of 61%, hence the cluster would not be unreasonable as a chance occurrence. In this case, no common risk factor or other reasonable explanation was found in spite of extensive investigation.

5 The aspect of time

Clusters of disease cases are often expected to be close in time as well as in space. This may be reasonable when the development of the disease or condition from exposure to the observed effect has a short duration, as would be expected for the cases of stillbirth or a contagious disease, for instance. However, in the case of cancer there may be a long delay, from several years up to several decades, between initiating cause and effect, and therefore one would not expect cases to cluster in time. In fact, in the cancer cluster discussed above, the cases stretched in time of diagnosis from 1980 to 1997, but still this cluster got attention because it was in a library in a small town, and with many long-term employees.

Expecting a cluster of cancer cases to be close in time is therefore generally unrealistic. If the cause that creates the cluster has an impact at an early stage of the cancer process, then there would be a large variation in time before a clinical cancer would arise. So closeness in time at initiation would not translate into closeness in time at the diagnosis of cancer. The only exception from this would be if the cause in question affected the cancer process at a very late stage. This would require a sufficient number of individuals in the community with latent tumours at such a late stage. In this case, it is disputable whether the putative cause can be said to cause the cancer, since latent tumours have to exist. But, conceivably, the cause may speed up the development of cancer. Even in this situation it might be difficult to observe a cluster effect because of variation in cancer growth rate between individual tumours.

Hence, a cluster of cancer cases that are close in time are not likely to have a common cause. This shows that the whole notion of cancer cluster as something that is also close in time is an unrealistic concept. This clearly points to the need for great caution in interpreting cancer clusters.

6 Variation in risk: Clustering as a signal

Although cancer cases with a common specific cause would not be expected to cluster in time, as explained above, there is another aspect to this. Often there is a variation in risk between communities; some have a higher risk than others. The simple fact which we shall explain is that the high risk communities have a higher probability of producing random clusters than

a low risk community. Hence, although the cluster will give a far too high incidence, it might still signal an actual increased risk. This is also a question of how sensitive the probability of clustering is to the underlying assumptions in the calculation.

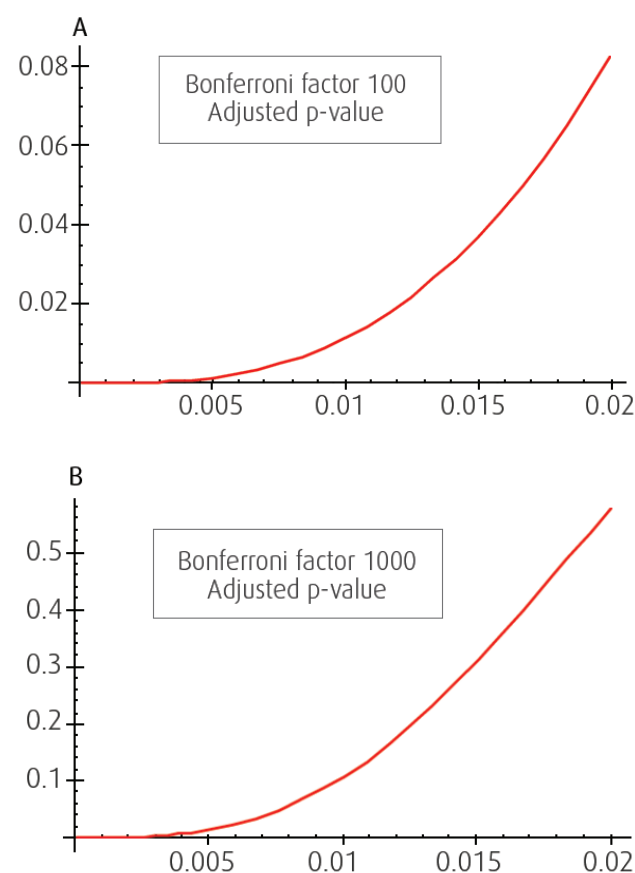


Figure 6 A and B. Kindergarten stillbirth example: Adjusted p-value of cluster as a function of background probability of stillbirth for two Bonferroni factors, 100 (A) and 1000 (B).

6.1 Stillbirth example

This is true not only for cancer but for any condition, and we shall demonstrate it for the kindergarten stillbirth example. We then make the same calculations as above (probability of 3 or more stillborn in 10 pregnancies), but for different values of the underlying probability of stillbirth. We also use two different values for the numbers of “similar units” to be corrected for, namely 1000 (as used previously) and 100. The results are given in Figure 6 where one sees that the adjusted p-value is strongly dependent on the underlying risk of stillbirth. Clusters of stillbirth are much more likely when the risk of stillbirth increases. This means that a cluster would be a much more probable event in a subgroup of the population with a generally higher risk. Hence one would think that a cluster could be seen as a signal of a generally increased risk.

6.2 Cancer cluster

Here we could also allow for the possibility of an increased risk. Possibly, the librarians as a group could have somewhat higher risk than the general population (this is here just used as a computational example with no implied suggestion of a general effect for this profession). Assume, for instance that they have twice the risk such that the expected number of cases rises from 0.73 to 1.46. Using the latter number as the expectation of a Poisson distribution yields a p-value (probability of 5 or more cases) of 0.0168 which is about 18 times the value computed previously in Section 4.2.

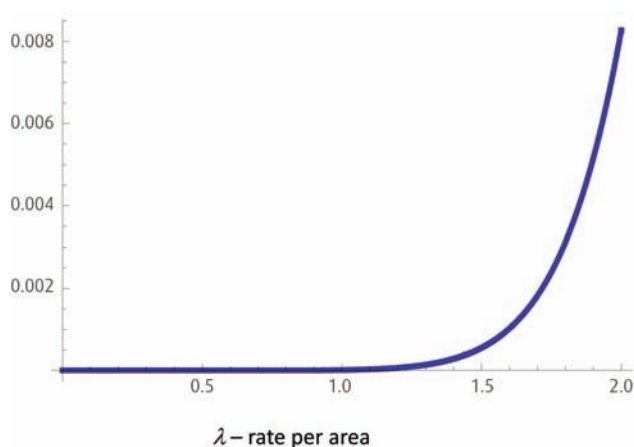


Figure 7. Probability of observing 11 cases or more in at least one out of 1000 units when expectation per unit, λ , varies from 0 to 2

If a cluster consists of more events (than the present five) the effect of increasing the underlying risk could be much stronger. Figure 7 shows an example where, using a Poisson distribution, we look at the probability of observing 11 cases or more in at least one out of 1000 units when expectation per unit, λ , varies from 0 to 2. There is a very strong increase after λ passes 1 and goes towards 2, in fact, the value at 2 is 823 times the value at 1. The reason for this can be seen from the Poisson formula $\lambda^x e^{-\lambda} / x!$. When x is (reasonably) large and λ is small, then the power λ^x implies a large effect of increasing λ .

Varying risk in populations is very common, and certainly a doubling in risk in some parts of the population compared to other parts is by no means extreme; see for example Sweeney & al (2008) regarding variation in breast cancer risk. We see here that such an increase can to a very large extent influence the likelihood of random clustering. Hence, although the cluster may be a random event and should not be interpreted as indicating a large local risk, it can still be a signal of a moderately increased general risk in parts of the population.

Notice that the cluster may itself be a very unlikely event, still it may contain important information. Using Bayes' law we may write

$$P(I|C) = \frac{P(I)P(C|I)}{P(C)}$$

where I denotes a raised general incidence and C denotes a cluster. Hence, the probability of a raised incidence given a cluster may be great even though the cluster may be an unlikely event because $P(C|I)$ is measured against $P(C)$.

7 Power to detect anomalies

We have so far focused on the interpretation of an observed apparent cluster. A closely related issue is to which extent an increased risk in a subgroup would be expected to result in a recognizable cluster. This is a question of statistical power. This calculation is only relevant for effects that arise over a short time, consider the discussion of time effects above.

Consider now the same Poisson process as in Section 3.2, but with the difference that one of the areas has a higher expectation $a\lambda$ where $a > 1$. Then the probability that the one with the higher expectation shall exhibit the largest number of occurrences can be derived from the previous formulas as:

$$P = \sum_{y=1}^{\infty} \frac{e^{-a\lambda} (a\lambda)^y}{y!} \left(\frac{\Gamma(y, \lambda)}{(y-1)!} \right)^{n-1}$$

To illustrate this we shall consider the following example: Assume that one area out of 10, or a 100, has a larger risk than the other areas. What is the probability that this shall stand out, that is being the one with the maximal number of cases? Consider two situations: (i) special area has twice the expected rate of the others; (ii) special area has five times the rate of others. The probability (power) that this area stands out from the others is demonstrated in Figures 8 and 9. In the case of merely a doubling of the risk the power is low, while it is relatively higher when the rate is multiplied by five in the special area. The figures also demonstrate the natural fact that correct discovery of the area with increased risk is easier when just a few areas are considered as compared to many (in the figures 10 and 100 areas are compared).

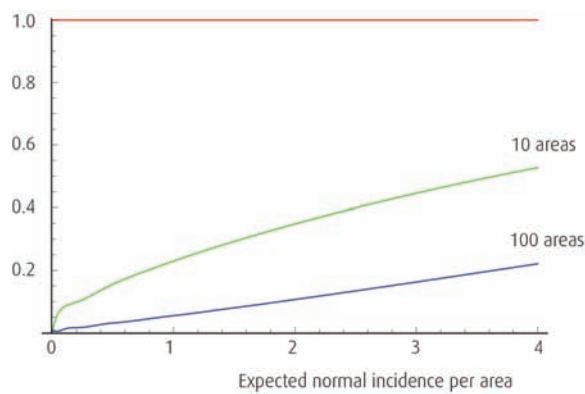


Figure 8. The probability for one area out of 10 or 100 to have the maximum number of cases, when the rate of this area is twice that of the others.

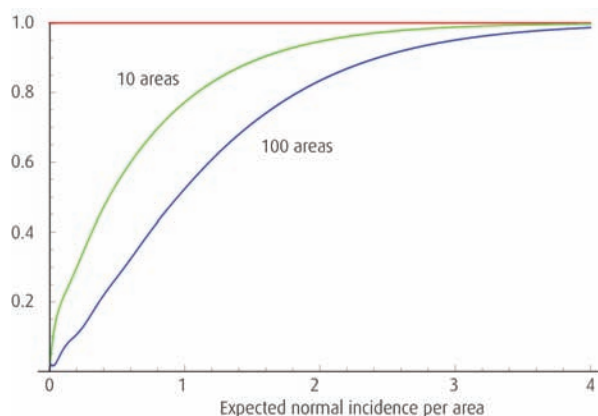


Figure 9. The probability for one area out of 10 or 100 to have the maximum number of cases, when the rate of this area is five times that of the others.

8 Final comments

Clusters of disease would often be expected to represent merely random variation. Whether they can be indicative of a real increase in disease risk, depends on whether the disease would develop fast (in which case they can), or slowly (in which case clustering in time is much less likely). In any case, a disease cluster could be a signal, not of a true cluster but of a generally increased risk level in a population. Clearly, clusters should be interpreted with caution.

References

- Aalen OO. Opphopning av sykdomstilfelle: noen statistiske betraktninger [Clustering of disease. Some statistical considerations]. *Tidsskr Nor Lægeforen* 1986; 106(9):761–763. [In Norwegian]
- Buchanan M. Statistics: conviction by numbers. *Nature* 2007; 445(7125):254–255. doi: <http://dx.doi.org/10.1038/445254a>
- Draper GJ, Stiller CA, Cartwright RA, Craft AW, Vincent TJ. Cancer in Cumbria and in the vicinity of the Sellafield nuclear installation, 1963–90. *BMJ* 1993; 306(6870):89–94.
- Gawande A. The Cancer-Cluster Myth. *The New Yorker* 1999 February 8: 34–37.
- Gavin AT, Catney D. Addressing a community's cancer cluster concerns. *Ulster Med J* 2006; 75(3):195–199.
- Hill AB. The environment and disease: association or causation? *Proc R Soc Med* 1965; **58**(5):295–300.
- Honjo S, Spencer DW, Farrington JW. Deep advective transport of lithogenic particles in Panama basin. *Science* 1982; 216(4545):516–518. doi: <http://dx.doi.org/10.1126/science.216.4545.516>
- Magerøy N, Møllerlækken OJ, Riise T, Koefoed V, Moen BE. A higher risk of congenital anomalies in the offspring of personnel who served aboard a Norwegian missile torpedo boat. *Occup Environ Med* 2006; 63(2):92–97. doi: <http://dx.doi.org/10.1136/oem.2005.021113>
- Palmer KT, Harris EC, Coggon D. Compensating occupationally related tenosynovitis and epicondylitis: a literature review. *Occup Med* 2007; 57(1):67–74. doi: <http://dx.doi.org/10.1093/occmed/kql127>
- Rothman KJ. A sobering start for the cluster busters' conference. *Am J Epidemiol* 1990; 132(1 Suppl):S6–S13.
- Rothman KJ, Greenland S, Lash TL (Eds). *Modern Epidemiology*. Third edition. Philadelphia (PA, USA): Lippincott Williams & Wilkins, 2008:1–851.
- Sweeney C, Baumgartner KB, Byers T, Giuliano AR, Herrick JS, Murtaugh MA, Slattery ML. Reproductive history in relation to breast cancer risk among Hispanic and non-Hispanic white women. *Cancer Causes Control* 2008; 19(4):391–401. doi: <http://dx.doi.org/10.1007/s10552-007-9098-1>
- The Lancet Oncology. Cancer clusters: how can we improve understanding? *Lancet Oncol* 2009; 10(12):1129. doi: [http://dx.doi.org/10.1016/S1470-2045\(09\)70355-9](http://dx.doi.org/10.1016/S1470-2045(09)70355-9)



Cluster inquiries, guidelines and lessons to learn

Tom K Grimsrud, Steinar Tretli, Tor Haldorsen
Cancer Registry of Norway, Department for research

Abstract

The attitude and approach towards perceived cancer clusters, and the scientific expectations from cluster investigation have changed materially during the last hundred years. In this article we refer some of the aggregated experience from the USA through recent decades, and we present selected examples of cluster inquiries addressed by the Cancer Registry of Norway. We also briefly refer to guidelines and factsheets provided by health authorities in the UK and the USA. Finally, we discuss the more general issue of resources, and we give a short presentation of emergency facilities for cluster investigation as they exist in the UK and Finland.

1 Introduction

During recent decades—in parallel with an increasing life expectancy and improvements in the health care system—we have seen a growing public awareness of health issues and environmental issues. Every month, the Cancer Registry of Norway receives questions from residents, health personnel, and media about local aggregation of cancer and potential environmental causes. Sometimes it is difficult to distinguish what comes first, the worry of a cluster, or an independent concern related to environmental exposure, such as electric high-voltage power lines, industrial pollution, radon, or garbage dumps. Common for these inquiries is the need for a qualified answer, and the opportunity it provides for health care workers to share their knowledge and become familiar with public concern.

One hundred years ago, doctors had virtually no knowledge of cancer causes. Encouraged by a Norwegian national committee for cancer research (Den norske komité for kræftforskning) general practitioners actively registered and reported cases in what they perceived as clustering of cancer (Gade, 1916). They discussed the possibility that some of the

cancers be caused by contagious agents (Hvoslef, 1903), see figure 1. The approach of these doctors is not unlike those we see from the public today, when people are faced with a perceived cluster.

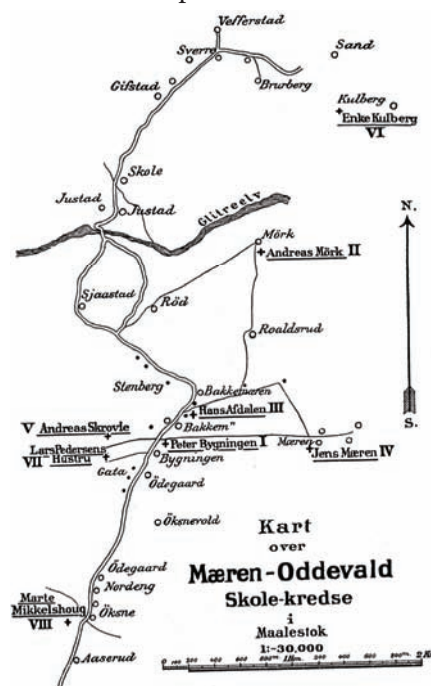


Figure 1. Investigation of a local “cancer epidemic”, printed 1903 in the Journal of the Norwegian Medical Association (Hvoslef, 1903). Note display of names and residences —contrasting present-day standards of personal secrecy.

The scientific and public health challenges linked to clustering of disease has been extensively discussed through the last three decades. In this paper, we will relate recent experience from handling of perceived cancer clusters, specifically from the USA, and we will refer to available guidelines and factsheets in the USA and the UK. Further, we will present some selected cluster inquiries that have been addressed by the Cancer Registry of Norway. Finally, from our neighbouring countries Finland and UK, we briefly describe systems for emergency handling of geographical clusters established during the 1980s and 1990s.

2 Reported experience from cluster inquiries in the USA

Towards the end of the 1980s, handling of cluster inquiries was recognised as a challenge needing operating procedures, and the US Centers for Disease Control and Prevention (CDC) arranged a national conference on clustering of health events with an international group of delegates in 1989. The proceedings from the meeting were published as a supplement to the American Journal of Epidemiology (Editorial Am J Epidemiol, 1990). A description of the CDC experience from 1961–1982, showed that the expectations by the public were largely shared by the health community, in the belief that investigation of cancer clusters would offer a clue towards unknown cancer causes (Caldwell, 1990). Initially, as referred from Norway (above), there was an anticipation of finding infectious causes to cancer clustering, which gradually were replaced by a suspicion of environmental contamination. However, the CDC experience from 108 cancer cluster investigations 1961–1982 did not provide a clear-cut explanation to any of the cancer clusters.

The CDC experience was in line with the provocative keynote presentation at the cluster meeting provided by the renowned epidemiologist KJ Rothman. He called it “a sobering start for the cluster busters’ conference” (Rothman, 1990), and his main point was to lower the scientific expectations to cluster investigations. Still, public concern can neither be denied nor rejected, and some of the less cited responses to Rothman’s “sobering start” may deserve attention. Neutra’s insightful “counterpoint” (Neutra, 1990) demonstrated that some cluster investigations indeed have been rewarding, although some researchers run the risk of ending up like Cervantes’s “Don Quichote” or Ibsen’s “An enemy of the people” (Norwegian: “En folkefiende”). The discussion

shows that a cluster inquiry can be discussed with two different approaches, the scientific one, and the public health approach. Obviously, the latter challenge cannot be met with advanced statistics alone.

Recently, another review was published of 567 perceived residential cancer cluster addressed by US federal and state agencies during the period 1990–2011 (Goodman & al, 2012). There was enough data to evaluate 428 of the reports, but for only 73 of the 428 (17%), the number of cancers exceeded the expected one and thus proved to be a confirmed cluster. For three of the clusters, there was some evidence of association with a suspected exposure, and for a single cluster, a clear cause was identified. Five of the investigations had been published in peer-reviewed literature.

The latter review by Goodman & al (2012) was funded by the Chlorine Chemistry Division of the American Chemistry Council—which remind us that industrial interests are not always compatible with those of the scientific community, nor with those of public health. Still, the described problems are in line with what should be expected from a scientific point of view: a mediocre scientific yield. For the sake of science, the advice may therefore be well-founded when the authors suggest (as did Rothman, two decades earlier (Rothman, 1990)) to give priority to larger studies with good exposure data to test more specific hypotheses. Nevertheless it may be necessary to address the question of perceived clusters.

3 Guidelines and information on cancer clusters

The US CDC provide useful general information about cancer clustering on the website of the National Center for Environmental Health (NCEH) (Centers for Disease Control and Prevention, 2012). The website provides definitions, advice on the interpretation of clusters, and information on the nature of cancer. Furthermore, the website offers link to the US Agency for Toxic Substances and Disease Registry (ATSDR) for guidelines in evaluation of perceived clusters. A stepwise approach is advised: case definition, case confirmation, population denominator, review of literature, exposure assessment, plausible hypotheses, and risk communication) (Agency for Toxic Substances and Disease Registry, 2000). A cluster inquiry may stop at any of these steps, for instance when an apparent cluster is not confirmed. The alleged cluster may often be based on a mix of different and non-related diagnoses, or enough information may be provided to show that the number

of observed cancers is not outside the range of the expected one.

Brief and well-written advice and descriptions of the cancer cluster issue are provided from several health authorities and textbooks. We would like to point at those of the US National Cancer Institute (National Cancer Institute, 2006), and the South West Public Health Observatory (SWPHO) in UK (South West Public Health Observatory, 2007). The latter is a National Health Service (NHS) organisation dedicated to help people access information about public health and improve the understanding of factors which influence health.



Figure 2. Map showing the localisation of some perceived clusters addressed by the Cancer Registry of Norway and briefly described in this article. Underlying map: Kartverket, Norwegian Mapping Authority, Norway

4 Perceived clusters addressed by the Cancer Registry of Norway

A cancer registry would be expected to provide useful data to the public and to a health worker facing a suspected or confirmed cancer cluster. Through the decades, a number of questions have reached the Cancer Registry from residents, health care workers, journalists, employees, and employers, see Figure 2 for the geographical distribution of selected inquiries presented in this article. In some cases, there is an underlying worry of a more or less poorly defined cancer hazard, and even other motives may be present, such as those of making a good story, economical interests in compensation, or a need to respond to an accusation of causing harm.

4.1 The Sømna story

An inquiry that really put clustering on the agenda was a suspected clustering of brain cancer in the municipality of Sømna in northern Norway. The story has been described (in Norwegian) by the former director of the Cancer Registry, Dr Frøydis Langmark (Langmark, 1994), and we shall refer only the main points here.

The Chernobyl disaster in April 1986 spread radioactive downfall in several directions, and some municipalities in northern Norway were among those perceived as more than average exposed. There was a marked public awareness, and countermeasures involved restrictions on intake of fish and meat from affected areas. In Sømna, cancer data were collected locally by a lay person, and the information was interpreted and disseminated as a dramatic cancer situation—even broadcasted on national television news on a Saturday evening (January 1993). No central expertise had been given the chance or time to evaluate the situation adequately.

There proved to be no increase in the total cancer incidence, but an update of the incidence at the Cancer Registry confirmed the suspected cluster, with six cases of cancer in the central nervous system against 1.6 expected from the background rates (standardised incidence ratio (SIR) = $6 / 1.6 = 3.8$; 95% confidence interval (CI) 1.4–8.3) for the six years combined following the Chernobyl disaster). However, the probability of observing one cluster of this size would be around 26 percent among 50 municipalities of the same size, a situation taken to be close to the real-life situation as 43 municipalities had been found to have a relatively high exposure. Three of the greatest challenges in the Sømna story were the delay in the production of quality-secured cancer data, the signs of media and population distrust in the health authorities, and the low interest among journalists in relating correct data to the public.

4.2 The oil refinery at Sola



Figure 3. Emissions from oil refineries have caused concern. A slight excess of total cancer was found among residents of Sola municipality, mainly involving cancers with no expected link to oil-refinery exposures. Picture from Mongstad, Hordaland county. Photo: Nina Aldin Thune

Between 1967 and 2000, an oil refinery was in operation in the municipality of Sola, situated between Stavanger and the North Sea, in south-western Norway. Former employees and local residents in the rural surroundings reported concern about the cancer incidence, and suspected that local pollution (smell, gas, and smoke) might have caused the cancers (figure 3). An epidemiological report was commissioned by the owner of the oil refinery, and the Cancer Registry studied the cancer incidence among employees, as well as the incidence among present and former residents in Sola. Estimates of residential exposure were provided by the Norwegian Institute for Air Research (NILU), and lists of residents and addresses in the municipality on the first of January for each year 1967–2000 were provided by the National Population Registry. Follow-up for cancer was obtained by linkage to national data in the Cancer Registry of Norway (Skog & al, 2009).

The oil refinery cohort counted only 450 individuals available for follow-up. A 10 percent excess was found for all cancers compared with the national population ($SIR = 1.1$; 95% CI 0.9–1.4; based on 67 cases), and no evidence emerged for an increase of cancers of the bone marrow, lymphomas, skin cancer, lung cancer, or cancer of the urinary bladder; cancers that might be of relevance in the oil industry.

The number of residents in Sola municipality rose from 8 000 to 19 000 during the period when the refinery was operated, and a total of nearly 44 000 people were available for a national follow-up after a minimum of 1 year's residence in the municipality. The entire resident cohort also showed a slight excess of total cancer ($SIR = 1.08$; 95% CI 1.0–1.1; based on 3018 cases), but the excess mainly involved cancers

with no expected link to oil-refinery exposures, such as breast cancer and ovarian cancer in women, prostate cancer in men, and skin cancers and bowel cancers in both sexes.

The estimated exposures from ambient air were quite low, and comparisons of groups with different cumulative exposure to benzene, polycyclic aromatic hydrocarbons (PAHs), or particles from the refinery displayed no clear pattern of dose-related incidence of relevant cancer forms. Separate analyses were conducted for women, and for children and adolescents in order to avoid confounding from occupational exposures. No specific data on potential confounding variables were available.

It should be underscored that the Sola oil refinery study was planned with 2.3 full-time equivalents (FTE) for skilled personnel over a period of two years, so the project was demanding and expensive.

4.3 Lung cancer in a municipality with iron and coke works



Figure 4. “Stoker in the State’s inferno; Now Gunnar is a cancer victim”. The presence of carcinogens may lead to speculations when colleagues or neighbours are diagnosed with cancer, ideas that are often amplified in the media. Printed with permission from Verdens gang, VG

During the 1950s through the 1980s there was a marked increase in the national lung cancer rate among Norwegian men. In the municipality of Rana—known for its coke works and iron smelter, both established after World War II—the rates for lung cancer and urinary bladder cancer among men surpassed the national rates in the 1980s showing no sign of levelling off. A reasonable question could be whether occupational exposures in the local industry could have contributed to the aberrant trends (figure 4). The two main employers were both run by the government, which secured funding for an

examination through a population-based case-control study. Cases of lung and urinary bladder cancer were identified in the Cancer Registry and controls were drawn among local residents according to the National Population Register.

Occupational history and smoking habits were collected by interviews with participants or next-of-kin, and work histories were supplied with personnel data from the coke works and iron smelter. The study identified an increased risk of lung cancer among pig iron smelter workers, possibly related to exposure to PAHs and asbestos. The investigation was published as a report and in a peer-reviewed journal (Grimsrud & al, 1998). Ironically, the local rate returned to the national level some years after the study was finished (figure 5), illustrating the problems of evaluating a recent shift in trends in a small area or small population.

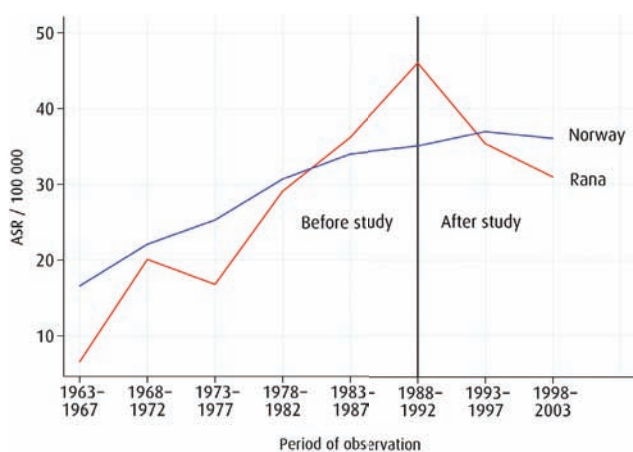


Figure 5. Age-adjusted rates (ASR, World standard) for lung cancer among men in Norway and in Rana municipality 1963–2003. The alarming rise until 1992 prompted a case-control study (Grimsrud & al, 1998), but later, the rates became non-excessive.

4.4 Breast cancer among teachers at a school

Several times, the Cancer Registry has been approached by teachers questioning the existence of an increased risk of breast cancer among their colleagues. Typically, they observe between 5 and 10 cases in the workforce consisting of 20 to 60 women through a decade or two. Below, we will refer some of the deliberations that have proved valuable in addressing these and similar cluster inquiries. Breast cancer is the most common cancer among women, and some 8 percent of women are expected to have a diagnosis of breast cancer before age 75. Among, say, 50 female teachers employed at a school through 15 years, of whom half have attained the age

of 75, we assume that about 2 cases of breast cancer would be reasonable to expect.

If we broaden our scope and assume that 2 breast cancers would be expected in each of 1000 schools during the same period, it can be shown mathematically that we should expect—by random, in line with a Poisson distribution—to observe a maximum of 8 cases in one school (see figure 5 in ref (Aalen & Tretli, 2012)). With another mathematical approach, it can be shown that the probability of observing 11 cases or more in one out of 1000 schools, given that 2 is the expected number, is approximately 0.008 (see figure 7 in ref (Aalen & Tretli, 2012)), that is, a quite unlikely situation. In Norway, there are nearly 3000 schools for children and/or adolescents (school year 1–10) (Statistics Norway, 2011), suggesting that the observation of 4 to 5 times the expected number of 2 cases would be likely to occur at a few schools.

The most important risk factors of breast cancer are being female, increasing age, not breastfeeding long term, current use of hormone replacement therapy, having no children or having no child before age 30, obesity (for post-menopausal women only), and a high consumption of alcohol. These risk factors may be more or less relevant for a group of teachers, but indeed, a large study of cancer risk by occupation (as noted in national censuses) demonstrated a 16 percent higher incidence of breast cancer among Norwegian female teachers (SIR = 1.16) compared to the general population (Pukkala & al, 2009). For the large group of economically inactive women, among them housewives, the breast cancer risk was in line with the population average (SIR = 0.99).

4.5 Increased cancer incidence in an area of a town (Mortensnes, Tromsø)

The Cancer Registry was contacted by the chief municipal medical officer in Tromsø, a town with 63 000 citizens in 2005, on the basis of repeated inquiries about a perceived increased cancer risk in a residential area (Mortensnes) of the town (figure 6). As routine data at the Cancer Registry are only reported at the municipality level, a simplified and coarse approach was explored in an attempt to give some idea of the size of the alleged cancer problem. Approximate numbers of person-years were derived from data on year of birth and sex for present-day citizens within the actual area postcode, and also from individuals who had died during the preceding 15 years with the same postcode as their last

address. Numbers of cancer cases diagnosed in the same period and carrying the area postcodes were provided from the Cancer Registry data base.



Figure 6. No evidence was found to suspect any strong cancer hazard in the Mortensnes area of Tromsø. Panoramic view of Tromsø from Fløya. Photo: Ragnilius

By this approach, we lost data for individuals moving out of the area in the observation period. The evaluation involved less than 3000 men and women, but it had some interest for a coarse evaluation of cancer incidence. The proportion of people above 60 years of age was higher than that of the national population, and the observed to expected number of all cancers taken together was slightly above 1, which is often seen in Norwegian urban populations. Cancer in the airways and digestive system was 30 percent above the national level, but within the limits of what could be expected by random variation. The evaluation gave no evidence to suspect any strong cancer hazard, and it had insufficient statistical power to identify weak effects (Grimsrud & Martinsen, 2006).

4.6 The Rosenberg laboratories

An indication of a cancer cluster emerged from the mid-1990s onwards among former students and colleagues at the governmental Norwegian University of Science and Technology (NTNU), Trondheim. Lymphohaematopoietic cancers were ascribed to work at the now demolished Rosenberg laboratories, and some patients were compensated by the Attorney General of Civil Affairs (Regjeringsadvokaten), despite substantial uncertainty whether there had been relevant exposures, or whether an increased risk did exist. No funding was obtained for an epidemiological investigation. The emergence of additional cases caused national media to roar in 2006, and an audit committee and a medical expert group were appointed by the government. The audit report was published as a Norwegian Official Report (NOU, 2007) and contained criticism against the university, and harsh criticism against the ministry.



Figure 7. Example of laboratory work (not the Rosenberg laboratories). Photo: Walkerma

In parallel, an epidemiological study was conducted by a team from the National Institute of Occupational Health, Trondheim University Hospital (St. Olavs Hospital), and the Cancer Registry of Norway. For the study, lists of students, fellows, and employees amounting to 7000 people were linked to data from the Cancer Registry. The risk of lymphohaematopoietic cancers in the entire group was close to that expected. However, the standardised incidence ratios in the subgroups of PhD students and participants in organic chemistry courses were increased, although based on small numbers (figure 7). The findings were in line with the assumption that PhD students and employees would experience higher laboratory exposures than undergraduate students. The study was published as a report (in Norwegian) and as a peer-reviewed article (Kristensen & al, 2007a; Kristensen & al, 2008).

A second investigation of students and personnel at the laboratories during the preceding 17 years (1960–1976, 900 individuals) confirmed the findings from the first report (Kristensen & al, 2007b). Despite the lack of good exposure data, the two reports left the impression of a causal link between long-term work at the Rosenberg laboratories and increased risk of lymphohaematopoietic cancers.

5 Cancer cluster resources

Mass media have the power to influence what is acceptable in society. Media pressure can be substantial and call for immediate action even in a question of chronic diseases. Some countries have established systems to address such situations.

5.1 Small area emergency units

In the UK, a Small Area Health Statistics Unit (SAHSU) was established 25 years ago, in 1987, in the wake of a cluster of leukaemia in children and young adults neighbouring the Sellafield nuclear plant (Small Area Health Statistics Unit, homepage). The SAHSU project started at London School of Hygiene and Tropical Medicine, and was later transferred to Imperial College (1995). An important purpose of the SAHSU unit is to serve health authorities and public with rapid evaluation of health risk in relation to environmental exposure based on routine statistics.

The SAHSU takes advantage of UK's mostly small-area postcodes combined with statistics on environment, health, and social characteristics. The data are used for research on potential associations between environment, socio-economic factors, and health, in the form of in-house or collaborative projects. The SAHSU has the advantage of serving a population of 60 million people living in a quite densely populated country.

Finland has another emergency system for small area analyses: Small Area Statistics on Health (SMASH). Data are organised according to a grid with 500 m times 500 m units, inclusive of population data on cancer incidence, residence, sex, age, and socio-economic status. Individual geographical information is provided at 2 points in time, and additionally, at time of a cancer diagnosis (Kokki & al, 2001). Different reference populations can be chosen according to the available parameters. The system does not reflect all dynamics linked to change of residence, and it has limited data on individual exposures, but it has been a useful tool for risk assessment and provides data for research. SMASH is involved in a number of international cooperative studies and networks.

5.2 Norwegian resources

There is no formalised system for the handling of cancer clusters in Norway, and no emergency unit exists. The Cancer Registry activity reported in this article represents only a fraction of the inquiries made

to the Registry every year, and additional numbers are probably handled by municipality health officers, health authorities, and other research institutions. The need for an emergency unit is, of course, greater for more acute forms of disease and death, such as those caused by contagious, infectious, or poisonous agents.

The Cancer Registry of Norway has been a central institution for aetiological cancer research in Norway. This experience has been useful in the evaluations of suspected cancer hazards. Traditionally, the Cancer Registry would prefer to link an investigation-prompted by a cluster or not-to an external health officer, public or occupational, that may be involved in long-term follow-up of residents or workers. The Registry publishes routine statistics of cancer incidence and survival every year, with recent and historical data on counties and the nation as a whole. More rarely, statistical reports have addressed regional or municipal rates, but the Registry regularly serves health officers and authorities with cancer data for a variety of purposes, among them for the evaluation of perceived cancer clusters. Personal secrecy can limit the degree of detail that can be provided when the number of cases is below 5 within a group. These challenges can often be overcome by merging of groups.

Strict rules for personal secrecy require licenses to be applied for, and regional committees for medical research must evaluate the project when sensitive data from different sources are to be linked and analysed. This procedure is time-consuming, and subject to application deadlines, committee deliberations and collection of additional information, and formal reporting. It is our experience that the generally low scientific value expected from small cluster investigations has the potential to hamper the approval by the ethical and legal boards, and to restrict the possibilities of quality control of a study. Countries with established emergency systems for cluster evaluations may have the advantage of a running license that allows for rapid investigations within defined guidelines.

Evaluation and communication of observational data on rates of rare diseases in small areas is a challenging task, but clearly an important public health issue. This activity has an educational side, and some skills in statistics, epidemiology, aetiological research, and public health issues are mandatory. The Norwegian population is well described, followed as it is by a National Population Register (identity, birth date, details on residence), and by Statistics Norway for a number of parameters (occupation, income, education).

In Norway, important health parameters at the county level, as well as general information on health issues are provided by authorities and non-governmental organisations. Information on cancer, cancer causes, and cancer occurrence may also be collected on the internet from foreign health authorities and cancer societies, such as those in the UK and the USA.

An epidemiological study with a unique design should be expected to last 2 years as a minimum after funding is granted, as it may include application and evaluation of licenses; appointing of qualified personnel; collection of background information, research data, and consent; linkage to existing registry files; quality control; statistical analyses; interpretation; and reporting. Qualified personnel is obviously needed, but not always readily available, and the price for a study is often linked to the quality of the work.

In a discussion of what would be a reasonable level of emergency preparedness to address a cancer cluster inquiry, one will have to weigh the costs against other funding of medical research. Furthermore,

one has to evaluate potential consequences from lack of timeliness when there is an urgent question of a suspected health hazards. Politicians are often faced with the choice between a higher degree of certainty resulting from properly conducted research with good design and appropriate dimensioning—a time and money-consuming task—and less certainty resulting from quick cluster assessments, which often remain inconclusive and much less penetrating. The educational sides of cancer clusters may be equally challenging for politicians, for the judicial system, and for media, as they are for the general public.

Abbreviations

SIR: Standardised incidence ratio; the observed number of cases divided by the expected number of cases, usually expected on the basis of age-, sex-, and period-specific rates in the national population

CI: Confidence interval, a range calculated to give a measure of how accurate your answer is, or of the uncertainty of the result.

References

- Aalen OO, Tretli. Can we draw causal conclusions from unexpected clusters of disease? In: Cancer in Norway 2010. Special issue: Clustering of cancer. Grimsrud TK (Ed). Oslo: Cancer Registry of Norway, 2012.
- Agency for Toxic Substances and Disease Registry. Disease Clusters: An overview, Evaluating a disease cluster. 2000. <http://www.atsdr.cdc.gov/csem/csem.asp?csem=20&po=5>
- Caldwell GG. Twenty-two years of cancer cluster investigations at the Centers for Disease Control. American Journal of Epidemiology. 1990;132(1 Suppl):S43-S47
- Centers for Disease Control and Prevention (CDC), National Center for Environmental Health. Cancer clusters. 2012. <http://www.cdc.gov/nceh/clusters/>
- Editorial Am J Epidemiol. Clustering of health events. Am J Epidemiol 1990; 132(1 Suppl):S1-S2.
- Gade FG. Undersøkelser over kræftsygdommene i Norge, paa grundlag av den officielle mortalitetsstatistik 1902-1911 samt det av den norske komité for kræftforskning samlede materiale 1908-1912 [Investigations on cancer diseases in Norway, based on official mortality statistics 1902-1911 and material collected by the Norwegian Committee for cancer research 1908-1911]. Kristiania (Oslo): Dybwad, 1916:p1-102. [In Norwegian] <http://archive.org/stream/skrifterutgitavv161chri#page/n5/mode/2up>
- Goodman M, Naiman JS, Goodman D, LaKind JS. Cancer clusters in the USA: what do the last twenty years of state and federal investigations tell us? Crit Rev Toxicol. 2012 Jul;42(6):474-90.
Doi: [10.1136/oem.58.5.31510.3109/10408444.2012.675315](https://doi.org/10.1136/oem.58.5.31510.3109/10408444.2012.675315)
- Grimsrud TK, Langseth H, Engeland A, Andersen A. Lung and bladder cancer in a Norwegian municipality with iron and steel producing industry: population based case-control studies. Occup Environ Med. 1998; 55(6):387-92. <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1757594/pdf/v055p00387.pdf>
- Grimsrud TK, Martinsen JL. Kreft på Mortensnes, Tromsø. Oslo: Cancer Registry of Norway, 2006: p1-4. [KRG-nettsider, skal legges ut igjen, kfr Tor Johnsen].
- Hvoslef A. En "kræftepidemi". (A "cancer epidemic"). Tidsskr Nor Laegeforen 1903; 23:689-696.
- Kokki E, Ranta J, Penttinen A, Pukkala E, Pekkanen J. Small area estimation of incidence of cancer around a known source of exposure with fine resolution data. Occup Environ Med. 2001 May; 58(5):315-20
doi: [10.1136/oem.58.5.315](https://doi.org/10.1136/oem.58.5.315)
<http://oem.bmj.com/content/58/5/315.full.pdf+html>
- Kristensen P, Hilt B, Svendsen K, Grimsrud TK. Forekomst av hematologisk kreftsykdom hos ansatte ved Rosenborglaboratoriene, NTNU. (Incidence of lymphohaematopoietic cancers in students and employees at the Rosenborg laboratories, NTNU.) Oslo: National Institute of Occupational Health, 2007a: 1-26. [In Norwegian]
http://bilder.bibits.no/stami/DIVrapp_utred/Rosenborgrapporter/Rosenborgrapport140207.pdf;
- Kristensen P, Hilt B, Svendsen K, Grimsrud TK. Incidence of lymphohaematopoietic cancers at a university laboratory: a cluster investigation. Eur J Epidemiol 2008; 23:11-15.
doi: <http://dx.doi.org/10.1007/s10654-007-9203-5>
- Kristensen P, Hilt B, Svendsen K, Grimsrud TK. Tilleggsrapport: Forekomst av hematologisk kreft ved Rosenborglaboratoriene, NTNU. (Incidence of lymphohaematopoietic cancers at the Rosenborg laboratories, NTNU.) Oslo: National Institute of Occupational Health, 2007b: 1-16. [In Norwegian]
http://bilder.bibits.no/stami/DIVrapp_utred/Rosenborgrapporter/Rosenborgnyrapport_endeligversjon.pdf

Langmark F. Clustering av krefttilfeller. Norsk epidemiologi 1994; 4(2):15–25.

<http://www.ntnu.no/ojs/index.php/norepid/article/view/191/169>

National Cancer Institute (NCI). Factsheet, Cancer Clusters. 2006.

<http://www.cancer.gov/cancertopics/factsheet/risk/clusters>

Neutra RR. Counterpoint from a cluster buster. Am J Epidemiol. 1990; 132(1):1–8.

NOU 2007:9 (Norwegian Official Reports 2007: 9). Rosenborgsaken. Oslo: Akademika, 2007: 1–79.

<http://www.regjeringen.no>

Pukkala E, Martinsen JI, Lynge E, Gunnarsdottir HK, Sparen P, Tryggvadottir L, Weiderpass E, Kjaerheim K. Occupation and cancer – follow-up of 15 million people in five Nordic countries. Acta Oncol. 2009; 48: 646–790. <http://astra.cancer.fi/NOCCA/Incidence/Results/Eng/08-Teachers.pdf>.

doi: <http://dx.doi.org/10.1080/02841860902913546>

Rothman KJ. A sobering start for the cluster busters' conference. Am J Epidemiol 1990; 132(1 Suppl):S6–S13.

Skog A, Martinsen JI, Aalen OO, Grimsrud TK. Solaprosjektet, Kartlegging av kreftforekomst blant tidligere ansatte ved A/S Norske Shells nedlagte oljeraffineri på Sola, og blant bosatte i området rundt raffineriet. (Cancer incidence among former employees at an oil refinery and among dwellers in the surrounding area.) Oslo: Kreftregisteret, 2009: 1–29. [In Norwegian]

http://www.kreftregisteret.no/Global/Publikasjoner%20og%20rapporter/solarapporten_web.pdf.

Small Area Health Statistics Unit. Homepage. <http://www.sahsu.org/>

South West Public Health Observatory (SWPHO). Factsheet No 18: Cancer Clusters. National Health Service: 2007: p1–8. <http://www.swpho.nhs.uk/resource/item.aspx?RID=9108>

Statistics Norway. Fakta om utdanning 2012 – nøkkeltall fra 2010. (Facts about education 2012 - key numbers from 2010.) Oslo: Statistics Norway, 2011: p1–35.

<http://www.ssb.no/emner/04/02/fakta/fakta2012.pdf>

Return address:
Kreftregisteret
P.O. box 5313 Majorstuen
N-0304 Oslo
Norway

Cancer Registry of Norway

Institute of Population-based Cancer Research

Postal address:
P.O. box 5313 Majorstuen
N-0304 Oslo
Norway

Office address:
Fr. Nansens vei 19, Oslo

Telephone: +47 22 45 13 00
Telefax: +47 22 45 13 70

E-mail: kreftregisteret@kreftregisteret.no
Internet: www.kreftregisteret.no

