[June 2017]

Factsheet:
Cancer Clusters

This factsheet is an updated version of a factsheet previously produced by the South West Public Health Observatory.

It has been produced by the UKIACR analyst group on behalf of the UKIACR. In particular the following have been involved:

Luke Hounsome	Public Health England (co-chair)
Ceri White	Welsh Cancer Intelligence and Surveillance Unit (co-chair)
Anna Gavin	Northern Ireland Cancer Registry
Hannah McConnell	Macmillan Cancer Support
Jana Witt	Cancer Research UK

What is a cancer cluster?

A cancer cluster occurs when more cases (of the same type or similar types) of cancer than expected are diagnosed in a group of people, geographic area and/or period of time.

When someone is diagnosed with cancer many people ask “what caused it?” This is especially true if several people they know are affected. Although great progress has been made over recent years into researching cancer risk factors and developing successful treatments, the question “what caused it?” remains very difficult to answer precisely.

Cancer is a complex disease with many different causes, and the reasons why it affects some people and not others are still poorly understood. Health professionals and researchers do not want to dismiss people’s concerns about cancer clusters but at the same time may be genuinely unable to provide the answers.

What causes cancer clusters?

Although most cancer clusters occur by chance (see p.5), it is not uncommon for people to be concerned that cancer clusters are caused by exposure to a cancer-causing agent in the environment. But real clusters that are proven to be associated with an environmental or occupational carcinogen are extremely rare.

Even if there are more people with one type of cancer in a community than might be expected, this does not necessarily mean that they were all caused by a cancer-causing agent in the environment.

Genetic makeup is an important risk factor for some cancers, which helps to explain why the same cancer may occur in different members of the same family.

Certain viruses have been identified as causal agents for some types of cancer such as primary liver cell cancer, Burkitt’s lymphoma and cervical cancer. Recent studies suggest that viral infections may also be associated with some childhood cancers, although in this case the exact types of virus are not known.

Lifetime risk of developing cancer

One in two people born after 1960 will develop cancer at some point before they reach age 85,.

The table shows the average lifetime risk for the most common cancers (figures from Cancer Research UK).

These risks may be higher or lower for individuals depending on lifestyle and family history.

Type	Men	Women
Lung	7.8%	6.2%
Bowel	7.3%	5.5%
Breast	0.1%	13%
Prostate	13.7%	-
Malignant melanoma	1.9%	1.8%
Non-Hodgkin Lymphoma	2.1%	1.7%

Causes of cancer

What is known about the causes of cancer?

There are over 200 different types of cancer that occur in different parts of the body. They have different causes, different symptoms and require different types of treatment. Risk factors are different for the different types of cancer; but for almost all of them the risk increases with age. The most significant and well-understood lifestyle risk factor associated with cancer is smoking. Other risk factors include an unhealthy diet, excess alcohol use, being obese or overweight, exposure to UV radiation, and a family history of cancer. It is thought that cancer is caused by complex interactions between these risk factors. Despite years of scientific research it is still very difficult to say for certain what the cause of cancer is in any individual case. The percentage of cancers that can be linked to factors in our environment beyond our personal control is probably small.

How are the causes of cancer found?

- Laboratory research is often necessary to understand the mechanisms that result in the growth of a cancer. For example, studying the physiological effects on the body of the presence or absence of a particular gene, or the effects of exposure to certain chemicals on cells.
- Another type of research, known as a prospective study, recruits a sample of healthy people and records extensive details about them, such as physiological attributes (weight, height etc.), fitness or lifestyle. These people are then monitored over several years in order to find out whether those that later develop cancer have any factors in common.
- Environmental risk factors can be studied properly if a *large-scale* community case-control study is feasible. This is only the case if:
    i. there is a definite environmental cancer-causing agent that can be measured and,
    ii. exposure levels to this agent in those who have the cancer can be compared with levels in those who do not, but the two groups are otherwise similar, for example in terms of age.
- There are several charitable organisations, research institutes and universities in the UK working together under the National Cancer Research Institute to conduct scientific research into the causes of cancer and effective treatments.

Cancer facts

- Cancer is the leading cause of death in the UK and Ireland.
- Around 1 in 2 people born after 1960 will be diagnosed with cancer at some point in their lives.
- The risk of developing cancer increases with age.
- Cancer incidence rates rise sharply among people over 45 years of age.
- In a population of 1,000 people with the average UK/Ireland age distribution, there would be, on average, 5 new cases of cancer each year. More cases would be expected in an elderly population.
- It is not unusual to know several people in a neighbourhood or workplace who have cancer.
- Cancer is not one disease it is a group of diseases characterised by uncontrolled growth and spread of abnormal cells. Different types of cancer have different causes and some are more common than others.
- While survival for many cancers remains low, recent advances in treatment mean that for some cancers, such as testicular cancer, malignant melanoma, and childhood leukaemia, the majority of patients now survive.
- The most significant lifestyle risk factor associated with cancer is smoking. Around 1/3 of cancer deaths are attributable to smoking. Other risk factors include an unhealthy diet, alcohol, exposure to UV radiation, and a family history of cancer. External environmental factors such as chemicals or radiation are thought to account for only 5–10% of cancers.
- Strong genetic predisposition accounts for around 5–10% of common cancers, such as breast and colorectal, predominantly affecting younger adults.
- Cancer does not develop immediately after contact with a cancer-causing agent. There is often a very long period, sometimes 20–30 years, between exposure and cancer diagnosis. Cancer patients who appear to be clustered may not all have lived in the same neighbourhood or worked in the same environment long enough for their cancers to share a common cause.

What can a cancer cluster tell us about the causes of cancer?

Investigations of isolated cancer clusters are unlikely to help us understand the causes of cancer in general. In the thousands of cluster investigations that have been carried out all over the world, the vast majority have not been able to find a cancer-causing agent in the environment that could plausibly be the cause of all the cancer cases in that cluster. Even where more cancers than expected are found and an environmental agent is suspected, it is not necessarily possible to conclude that a cause and effect relationship exists. Cancers take a long time to develop and exposures, if any, which led to the disease could have occurred years, even decades, ago.

Determining any environmental contribution to the cancer in a particular neighbourhood, where new people have moved in, and many have left, is difficult. The existence of a cluster may generate ideas about what caused it, but far more extensive research would be needed to test these ideas. Questions that need to be asked include:

- Have all the cancer cases in the cluster been exposed to the suspected cancer-causing agent for long enough?
- Is there a proven or plausible mechanism by which this agent causes cancer?
- Can exposure be measured for everyone in the cluster?
- Are there any factors which need to be ruled out, such as smoking or genetics?

A study looking back at patients' lifestyles is usually not possible as many may have died or are too unwell to take part in research.

Mesothelioma – an example of a real cluster

Mesothelioma is a cancer of the membranes that surround the lungs and abdominal cavity. It was practically unknown before the 20th century. In the 1950s, health workers in South Africa's North Western Cape began to notice high numbers of a strange cancer in people living near the asbestos mines. In 1960, a study of 33 patients revealed that 32 had been exposed to crocidolite (blue asbestos). Researchers proposed that mesothelioma would be found among workers of any industry using blue asbestos and further studies confirmed that this was, in fact, the case. As the dangers to health became apparent, most types of asbestos were banned in the UK.

Several factors were important in proving the link between the original mesothelioma cluster and asbestos. Most important was the fact that mesothelioma was always linked to cases of industrial use of blue asbestos. Other factors included the presence of asbestos fibres inside tumour tissue.

Although asbestos has been banned in the UK for many years, the number of new mesothelioma cases per year is still rising due to historical exposures, as the disease can take decades to develop.

Investigating clusters

What happens when a suspected cluster is reported?

The first stage is to find out about concerns in the local community and to gather some basic facts about people affected, including:

- the type(s) of cancer;
- the number of people reported with the same type of cancer;
- the ages of the people affected;
- the time period in which they became ill;
- the location;
- whether a specific agent in the environment is suspected, and
- how long this agent has been around.

The local Director of Public Health can then decide on whether and how this will be investigated. This may include an analysis of routine data from the cancer registry, or a more detailed investigation involving environmental health officers and others.

In most cases a large-scale study is found not to be necessary.

What steps are involved in the analysis of routine data?

If the local Director of public health asks the cancer registry for their assistance, they will aim to follow the UKIACR guidance for cancer registries on using their data to look at cancer clusters.

http://www.ukiacr.org/publication/investigating-and-analysing-small-area-cancer-clusters

Do the results indicate a real cluster?

The following criteria are usually used to determine whether a cluster is likely to be real and whether further study is feasible:

- a measurable cancer-causing agent is present in the environment and this agent could have caused the type(s) of cancer seen;
- the standardised incidence ratio (SIR) shows that the number of new cancer cases is several times higher than expected and that this is unlikely to be due to chance.

If these criteria are not met, then further local study will be unable to answer the question as to what caused the cancer cases in question. The investigation cannot completely rule out the possibility that the cases are linked – only that it is unlikely that a common link exists. Most investigations find that it is likely that the cluster has occurred due to chance. The raised number of cancer cases may be temporary, and disappear in later years. The most usual outcome of a cluster investigation is that the number of cancer cases in the area of concern will be monitored closely in the future.

Texas Sharpshooter

The ‘Texas Sharpshooter’ shoots at the side of a barn and then draws a bulls-eye around the best cluster of bullet holes. In the same way, clusters of disease are often isolated from their context. People tend to notice cases first, for example four cancer cases on the same street, and then define the population to be studied as only those in that street. Investigators can’t be sure whether they are creating a cluster by drawing lines around a few unconnected cases.

It is very important to ask questions such as: “If there is something affecting this street, what other neighbouring areas would it also affect?” and “What is the smallest population that should be studied?”

How can clusters be due to chance?

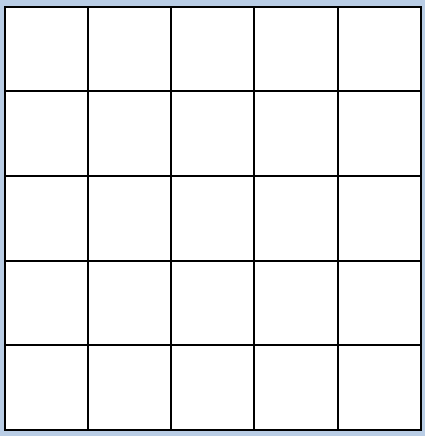
- It is thought that cancers develop because of a complex interaction between our genes, our living environment, and our lifestyles. Only for very few cancers can a single factor be identified as the main cause, such as smoking and lung cancer. For many other common cancers, such as breast and bowel cancers, it is not possible to find a single cause. It is likely that a combination of factors is involved and that there are chance interactions between these factors that lead to the development of a cancer.

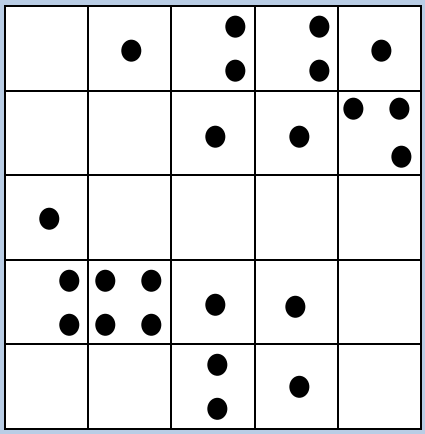
- Random events do not occur regularly. For example, in the National Lottery, the balls are drawn at random and each ball has an equal chance of being drawn. However, by the start of 2004, ball number 8 had featured in three times as many draws as ball number 20. This does not mean that the lottery is rigged. A feature of randomness is that some events will occur less frequently than average and some will occur more frequently than average. In the same way, some areas have a higher number of cancers than average and other areas have a lower number of cancers than average.

- Imagine that a populated area is divided up into a grid of 25 squares with 200 people living in each square. According to the average number of new cancer cases in England, each square would expect to see one new case of cancer in a year. Over the whole grid there would be 25 new cases each year expected.

- In a random distribution, you would not necessarily see one case in each square. The grid below shows an actual result of generating 25 random numbers between 1 and 25 and putting a dot in the grid corresponding to each number. There is one square with four dots and several squares with no dots. In the square with four dots, the cancer incidence is four times that which would be expected, according to the average.

- Even with a very rare disease, there is always a possibility that, just by coincidence, somewhere, sometime, several cases will arise in people who live near one another. This makes it very difficult to distinguish between clusters of diseases that have a common cause and clusters that are due to chance alone.





Cancer registries

Functions of a cancer registry

Cancer registries are statutory bodies which collect data on all new cases of cancer diagnosed within their region. Cancer registry data are used to monitor patterns of cancer, especially over time and across geographical areas. The registry provides information to local agencies and this is used to evaluate improvements in health and equality. The data may also point out possible public health problems and highlight areas for action. Cancer registries support cancer services in the planning and provision of care for their patients. Registries also contribute data to research organisations for their research into the causes of cancer and the best ways of treating it.

Using cancer registry data to analyse suspected cancer clusters

Advantages

- The data are population-based, this means they capture all cases of cancer within the population.
- Long-term time trends are available.
- Data are available on all different cancer sites and types.
- Cancer registrations include demographic information such as age and sex.
- The number of new cancer cases in small populations can be compared to larger areas.
- Enable accurate, unbiased study, independent of any local surveys.

Disadvantages

- The cancer registry data do not routinely include information on lifestyle factors such as smoking or diet, which may affect cancer risk.
- The registry holds little or no information on how long people have lived in an area or where they lived before. For many cancers it can take years from exposure to a cancer-causing agent to development of the cancer. Prolonged exposure over many years may be necessary to cause cancer.
- The registry has very little information on exposure to known cancer-causing agents at work, such as asbestos, chemical dyes or radioactivity – only what is mentioned on death certificates.
- Data are obtained from many different sources and it takes time to collate and check information about each patient. The most recent complete registration year is usually around one year out of date. It is therefore difficult, although not impossible, to investigate suspected clusters that have occurred very recently.

What cancer registries *can* do

- Analyse data to establish whether the area of concern has a higher number of cancers that is not due to chance
- Highlight cases where there is evidence of a risk to public health and inform the appropriate agencies.
- Work with other agencies to identify risks.
- Provide data to organisations to support research into the causes of cancer and new treatments.
- Keep up to date with new research into the causes of cancer to inform future cluster investigations.
- Look at patterns of new cancer cases on a large scale.

What cancer registries *cannot* do

- Get samples from the environment and test them to analyse potential cancer-causing agents. This is often the responsibility of the local environmental health department or other organisations.
- Undertake large-scale prospective or case-control research studies. Such a study would require the participation of hundreds, if not thousands, of subjects and would require the resources of a big research organisation.
- Therefore, provide conclusive proof of a link between environmental factors and cancer.
- Establish the causes of cancer through the analysis of registry data alone.

More about cancer registries

More information about the cancer registries is available online:

http://www.ncr.nhs.uk/patientinfo/

http://www.wcisu.wales.nhs.uk/home

http://www.isdscotland.org/Health-Topics/Cancer/Scottish-Cancer-Registry/

https://www.qub.ac.uk/research-centres/nicr/

http://www.ncri.ie/

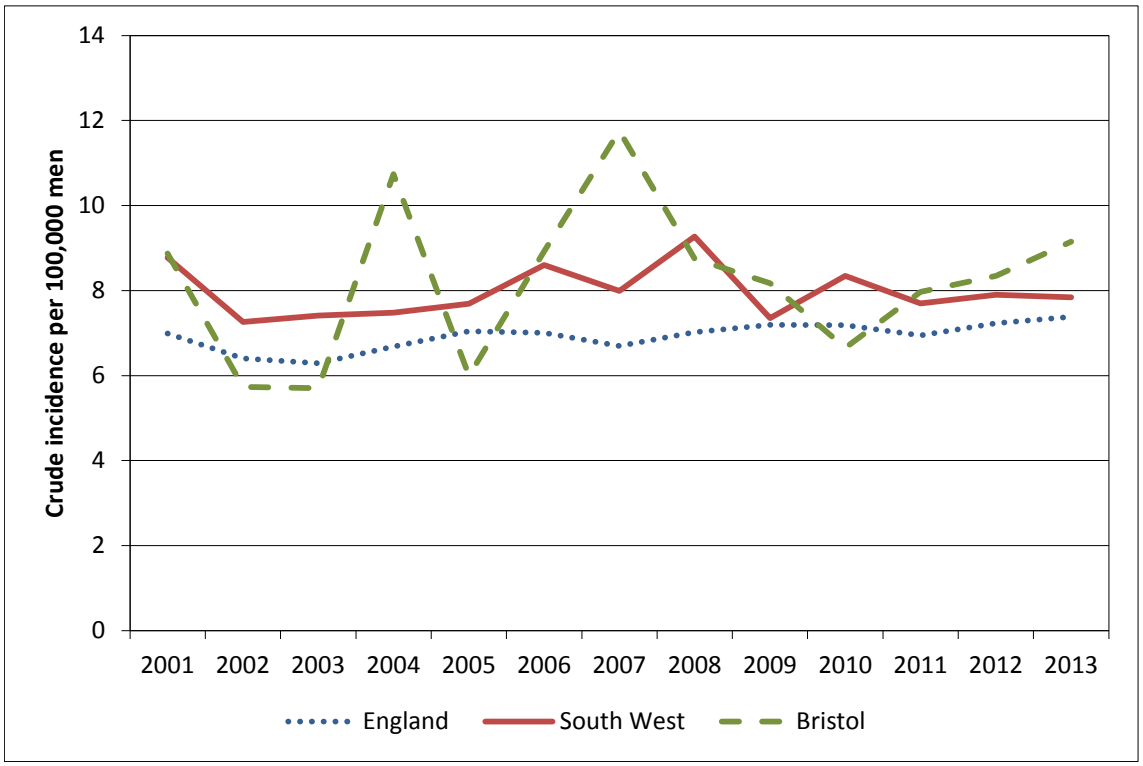
Analysis of small numbers of cases

It is often difficult to interpret and present the results of studies that look at very small numbers of cases or small populations. Some of the reasons why this is are explained below:

Annual fluctuations

In small populations, the number of new cases of disease over time (called the 'incidence rate') is likely to vary a lot, especially for rare diseases. The graph below shows the incidence rates of testicular cancer in three differently sized area. The South West region has a male population of around 2.7 million. This is compared to England with a male population of around 52 million and Bristol city with 220,000.

The graph below illustrates how incidence rates of a rare disease such as testicular cancer show greater variations from year to year in the smallest area. In Bristol there appear to be sudden and dramatic increases in the incidence rate of testicular cancer in some years but, in fact, the average incidence trend over the period is the same in all three areas.



Coverage

When investigating a reported cluster, it is important to ensure that all cases of cancer in the area concerned are identified and that the data are carefully checked and validated. Missing just one case, or counting a case twice, can have a huge impact on the statistical significance of the results, as illustrated by the examples below (remember that a p value of less than 0.05 is the usual level at which the results are said to be statistically significant, which means they're unlikely to be due to chance).

Example 1

The expected number of cancer cases in a population is five. There are eight actual cases (in which the p value would be 0.07, which means not significant), but one case has been recorded twice and 9 cases are counted. The p value is interpreted as 0.04, which means it is statistically significant and so not thought to be by chance.

Example 2

The expected number of cancer cases in a population is five. There are nine actual cases (in which the p value would be 0.04, which means is statistically significant) but one case has been overlooked and 8 cases are counted. The p value is interpreted as 0.07, which is not statistically significant and so thought to be by chance.

Confidentiality

Cancer registries are not permitted to release any data that could lead to the identification of a patient (except under strict confidentiality rules) or to publish any data that could be used to identify a patient. Small numbers of cases in a small area are regarded as identifiable information. The table below gives an imaginary example of the number of cases of cancer in men in a small village in England.

Age group	0-44	45-64	65-74	75-84	85-89	90+
Number of cases	5	10	15	15	5	1

If there is only one male aged over 90 living in the village then the table is directly disclosing confidential medical information about this person. The UKIACR recommends that no statistics should be published that are based on a count of less than 5. This applies to straightforward counts as well incidence rates and ratios, where the actual number of cases can be worked out if the population is known.

As a further precaution, cancer registries do not release information based on a population of less than 1,000, unless data confidentiality agreements have been signed and the data have been anonymised in some other way.