Standard Operating Procedure:

Investigating and Analysing Small-Area Cancer Clusters

This Standard Operating Procedure (SOP) has been developed by a subgroup of the UK and Ireland Association of Cancer Registries (UKIACR) Analysis Group, with important contributions from Luke Hounsome[1] (lead), Jason Poole[2], Matthew Barclay[2], Sean McPhail[3], Ceri White[4], Diane Edwards[5], Sandra Deady[6]. The Analysis Group will maintain and update this document.

[1]Knowledge & Intelligence Team South West, Public Health England (PHE); [2]Knowledge & Intelligence Team East Midlands, Public Health England (PHE); [3]National Cancer Intelligence Network, Public Health England (PHE); [4]Welsh Cancer Intelligence and Surveillance Unit, Health Intelligence Division, Public Health Wales; [5]Knowledge & Intelligence Team West Midlands, Public Health England (PHE); [6]National Cancer Registry Ireland.

Further information on the content of this SOP is available from your regional or national cancer/public health intelligence lead. Their contact details are available from the UKIACR website: http://www.ukiacr.org/ .

Background

The organisations which analyse cancer registration data in the UK and Ireland are members of the United Kingdom and Ireland Association of Cancer Registries (UKIACR). These organisations often receive requests about the incidence of cancer in small areas, frequently referred to as cancer 'clusters'. These can cause considerable concern for the public and other stakeholders, and often take considerable time to analyse and report on. The response needs to be carefully considered or analysis may need to be repeated, or worse there may be accusations of 'manipulation' of the data.

It was agreed that there should be a standard operating procedure (SOP) in place for dealing with these requests, to serve two main purposes:

- to ensure that an appropriate response is given to these requests; which deals with the requestor's concerns but is reasonable in terms of resource use
- to help analysts who are unfamiliar with these type of investigations to be aware of issues they need to consider, and undertake a useful analysis

Those cancer analysts who work for Public Health England (PHE) also have guidance in the forthcoming 'Guidance for investigating non-infectious disease clusters' document produced by PHE.

This document is designed for analysts who are familiar with cancer registration data. It is not intended for a member of the public. There is a factsheet on cancer clusters which serves this purpose (http://www.swpho.nhs.uk/resource/item.aspx?RID=9108), and is due to be updated by the UKIACR analysis subgroup.

Defining a cancer cluster

A cancer cluster is an unusual number of cases of the same type or similar types of cancer occurring during a specific time period among people living in a defined geographical area or people with the same working environment.

This definition notes that the cases should be similar (to allow for a single cause) and in a defined area (to allow for a point source). The random nature of biological processes and the laws of statistics mean that at any given point in time some areas will have a higher incidence of disease than others, and hence some criteria are required to try and reduce the possibility that random effects are responsible for the number of observed cases.

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

•   the Standardised Incidence Ratio (SIR) is statistically significant
•   is a measurable point source of carcinogen suspected or been identified?
•   could this carcinogen have caused the type(s) of cancer seen

As well as being statistically significant a number of sources have attempted to define a value of the SIR which is likely to indicate a localised source of carcinogen. These tend to focus on the range of 2-5 times the expected number of cases. However this will need to be put into context on each occasion depending on the cancer type, feasibility of a point source and local factors. For example, an SIR for all cancers of 2 in a deprived area is perhaps not unlikely; whereas an SIR of 2 for bladder cancers near an industrial heavy-metals site could warrant further investigation.

These criteria dictate that initial investigations should be at a fairly high level with the intention of determining if further work is needed.

Sources of requests

Requests for data on cancer in small areas, or potential 'clusters' can come to UKIACR organisations through several routes:
•   the public
•   GPs
•   Health organisation (e.g. Clinical Commissioning Groups, Health Boards, NHS)
•   local authorities/councils

- elected officials (MPs/MSPs/TDs/AMs/MLAs/councillors etc.)
- campaign groups

The majority come directly from the public, or by the public contacting a third party e.g their GP or elected representative. Some arise from campaign groups, particularly in respect to nuclear power, mobile phone masts, and fluoridation.

These different audiences will require different approaches, taking into consideration the following factors:
- the concepts of random variation and inherent instability of small numbers may have to be communicated to a lay audience
- there may be differences in the geographical area the requestor would like analysed, and the geographical area we can analyse. This will need to be communicated
- there is poor knowledge of the type and amount of data available to the UKIACR organisations around cancer; even from other health professionals
- health professionals may wish to have data which we would not release to the public, e.g. incidence data for potentially identifiable age/sex groups. This may mean we have two versions of the data in circulation which will require comprehensive version control.

Receiving the request and managing the initial response

If a request is received by the cancer registration service or the information analyst then the name and contact details of the requestor should be noted and the information passed to the relevant Director of Public Health (DPH) for the local area. The DPHs are located in:

- Health Boards in Wales
- County Councils, Unitary Authorities or London Boroughs in England
- NHS Boards in Scotland
- the Public Health Agency in Northern Ireland
- former Health Boards in Ireland

When the information is passed to the DPH it should be agreed that they or an individual nominated by them will lead and respond to this request.

Those analysts working for PHE should also notify the relevant PHE Centre that they have received the request and summarise any planned work/response.

It is essential to build up a relationship that will facilitate information sharing between the public health professionals and the requestor; this may be best left to the DPH in difficult requests but will be dependent on their availability. Ideally this should be done via a face to face meeting or at least a telephone conversation in the first instance. Do not at any time dismiss the requestor's concerns. During these initial conversations it can also be beneficial to provide some background information to the requestor in terms of information bulletins, cancer statistics or how random clustering can occur (See fact sheet: http://www.swpho.nhs.uk/resource/item.aspx?RID=9108).

At this stage the key points to cover are:

- contact information for the requestor and for the liason person in public health e.g DPH (so that the requestor can get in contact with the designated individual at any time)
- acknowledge the concerns of the requestor
- ask for any additional information that they have with regards the number of people affected. It may be necessary to take detailed demographic and clinical details for the cases to verify the cases recalled were diagnosed as cancers
- tell them the next steps and inform them when they can expect to be contacted again
- a summary of the conversation should be provided to the requestor including any timescales that were agreed

Do not commit resources and time above and beyond what is possible and do not promise to undertake detailed study or analysis at this stage.

It can be helpful at this point to verify the reported cases and exclude duplicates. Additionally for the geographical area under investigation, there is a need to investigate and uncover all cases which might be registered but that the concerned public or individual might not be aware of. This will increase the reliability and credibility of the work.

The information provided by the requestor should be treated as confidential and reviewed to consider what further investigations are warranted. Some criteria that will help guide this decision include:

- which cancer sites have been identified?
- are there specific concerns related to either sex?
- what time-period are the concerns for?
- is there a specific geographical area of concern?
- are the different types of cancers not known to be related to one another?
- are there few cases of a common cancer (e.g. lung, breast, prostate) or of rare cancers (e.g. liver, bone, brain)?
- are there cancers in family members who are genetically related?
- are the cancer cases among persons who lived in the same geographic location in the same timeframe?
- do we know anything about a trigger factor or latency in this/these cancers?
- how many cancer cases have been identified?
- identification of a hazard or environmental cause
- identification of an occupational or community setting
- have cancer deaths and new cases been counted together in the public perception of cancer?

A decision should be made as to whether this perceived cluster has any public health significance. If it is unclear whether the cluster is of public health significance then analysis should be undertaken where both parties agree the analyses to be carried out; again this is likely to be a decision left to the relevant DPH. It is often best to agree analyses and deliver the results via a face to face meeting or over the telephone, including when the perceived cluster is clearly not of public health significance and no further analysis is warranted.

Defining the analysis

Initial analysis should be based on published incidence rates for administrative and health areas in the location identified by the enquirer. This assumes the perceived cluster has been assessed as being of potential public health significance and therefore worthy of further investigation (as above).

Case confirmation

When the cluster involves cancer, it is useful to confirm that the enquirer's facts are correct before agreeing to do more detailed analyses. If the DPH has collected demographic and clinical details of the cases known to the enquirer, the cancer information team can confirm the epidemiology of the cases to inform on the extent of our analyses. It is important not to reveal any new details obtained from case verification to the enquirer.

The demographic details that should be collected for each named individual are:
•	their full name (and any previous names),
•	their date of birth (age if this is not available),
•	their NHS number or other unique identifier,
•	their current address and any previous address (with postcode) ,
•	the date from which the person has been living in the neighbourhood ,
•	information on current and/or previous occupation may be important for some cancer sites (e.g. bladder, leukaemia),
•	it can be useful to verify what screening activity has been taking place in the area if the alleged cluster involves breast, cervical or colorectal cancer cases.

Clinical details of each individual should include:
•	the cancer site and histology,
•	their NHS number or other unique identifier,
•	the hospital(s) of diagnosis and treatment,
•	for cancers such as breast and prostate cancer, details of treatment such as HRT or information regarding PSA testing may also be important.

Cancer types

In defining the study we need to be clear which cancers should be included and excluded from the study e.g if it is colorectal cancers that appear to be on the increase do we include all parts of the colorectal canal or specific parts?

Sex/age groups - Specifying the sex/age groups of concern may be important if the enquirer thinks there is an increase in incidence amongst the young or elderly, males, females or both. Agree which groups should be investigated before analysis starts.

Time frame – the enquirer may think the increase in cancers has been in last 3 months, this may be impossible for us to prove but we can look at a time series to identify any peaks and troughs in rates and compare these with 3 year rolling averages – suggesting there is random short term variation.

Occupational groups – it could be a group of employees in one industry that has raised concerns. If they live locally the study could initially rest on published incidence figures in administrative and health geographies. If not there may be a case for asking employees consent to share data e.g. home postcodes, and details of time in employment, to check if there are enough employees to produce rates for the cohort employed in the industry. This should be agreed in advance and in agreement with the Health and Safety Executive (HSE) or similar body.

Geography

We should assess cancer incidence for an area relevant to the concerns of the requestor. We often find that typical administrative or health areas cover too large an area to be useful, but to make simple, valid comparisons over time we want to use areas of consistent size and with accurate, granular population data.

If the concerns centre around a particular point source, we may find it useful to ask environmental health teams (e.g. CRCE within PHE) for advice. They may:
• be able to identify areas close to the point source we should include in our investigation;
• know of other possible sources of carcinogens in the area; and/or
• be able to provide evidence that the point source is not likely to be causing cancers in the local population.

We typically use census geographies to build some sensible radius around the point source or covering the area of concern. This may lead to 'odd'-shaped areas or to us investigating cancer incidence in a much wider area than the requestor thinks is appropriate. However, we do not have reliable population data for smaller areas which makes any assessment of cancer incidence questionable.

It can be difficult to know which area(s) to analyse, and this will be based in part on what the requestor describes. We have to be aware though that the requestor may have selected the area based on knowledge of cases, rather than an objective selection. This is the so-called 'Texas Sharp-Shooter' problem. We should ask: does the area selected fit with any hypothesis the requestor has

put forward? If there are requests to amend the area studied, be aware that the requestor may be 'fishing' for a result.

Census geographies are unlikely to be well known to the requestor and are not necessarily intuitive to work with. We usually find it helpful to use Geographic Information Systems – or just a large-scale map – to help choose areas to include in the analysis, and should generally present a map with the results so it is clear which area we have investigated.

We will also need to choose an area for comparison of rates. This should generally be large enough that cancer incidence rates are stable. We often use national cancer incidence rates or incidence rates for the county or region the perceived cluster is in.

We may find it useful to consider incidence in a comparable small area, similar to the area of concern. This can make communicating the results more straightforward.

Time periods

We usually find it helpful to look at longer time periods, where possible, because this allows for a more reliable assessment of cancer incidence. We can also use this to show how incidence varies from year-to-year, which may be an easy way to describe the amount of natural variation we would expect to see. However, we need to be aware of the concerns of the requestor, and make sure our time periods are appropriate and helpful in assessing them. For example, if there was concern about a waste processing plant which opened five years ago, we would probably want to consider the five years before the plant opened and the five years since the plant opened so we can clearly see how cancer incidence has changed. The change in population structure in this sort of time period can be significant.

Occasionally, the requestor is interested in incidence in a very short period of time, perhaps believing there has been a sudden increase in cancers over the past three months. This is generally impossible for us to reliably investigate due to small numbers and timeliness of cancer registration, so we should always look at longer time periods.

Ages, sexes and other variables

We need to agree the sex and age groups we are investigating before we begin analysis. If the requestor is concerned about cancer incidence in elderly men, we should restrict our analysis to this group and should not include cancers in, say, women of working age. For rarer cancers, we may not be able to restrict our analysis to just one age or sex grouping.

We should also consider whether we should adjust for factors such as deprivation. We know that many cancers are more common in deprived areas, and that certain cancers are more common in affluent areas. This is usually because of various lifestyle factors which are associated with deprivation. For cancers with a known strong deprivation link, it is often appropriate to adjust for deprivation in our analysis. As with all other choices about the analysis, this should be agreed before we begin analysing the data.

Contentious issues

There are certain factors relating to the process of investigating small-area cancer statistics which are often viewed by the public as contentious and will need to be handled with sensitivity.

The most common question is why we cannot provide more recent numbers than we have available especially if it is felt the increased numbers of cancers have only just happened. It is important to emphasise the fact that the data available to UKIACR is world-class, but that it requires time to collect, process and quality assure the data properly. There are delays between diagnostic testing and treatment so that data may not be immediately transferred to the registry nor confirmed with other demographic details for immediate use in analysis.

The request will often be for one year, or a short time-period where cases have 'spiked'. This may well be at odds with our need to aggregate several years for statistical robustness. This ties in with the note above on communicating the complexities of random processes.

Concerns regarding high cancer rates may be focussed around a place of work, or education. (See also note above on 'custom' geographical areas.) This is nearly impossible for us to do with routine data. In this situation we must offer to analyse a comparable area, whilst acknowledging it is not exactly what was asked for. It may be helpful to refer to the Health and Safety Executive (or equivalent) for known industrially linked cancers.

Benign/in-situu/recurrent/metastatic cancers are all counted as cases in the public eye. It is important to clarify exactly what the request relates to, and communicate what we cannot include e.g. metastases. We should acknowledge that these manifestations of disease, while not independent primary cancers, are still significant life events (see Standard text section).

The incubation period for cancer varies according to whether the carcinogen affects early steps in tumour development (in which case risks may not increase for 10 years later or more) or later steps. However in general cancers have long average incubation periods (Armenian, 1983)

There is both left censoring (people who have developed cancer after moving into an area but who were exposed elsewhere), and right censoring (people who develop cancer after having moved away from the area where they were exposed).

Data available

We are fortunate to have postcode-level cancer incidence and mortality data and in most cases full postal address available; although we do not have access to individual lifestyle data. The limiting factor for our analyses is often therefore the level at which we can access population data through time. This means that analysis will generally have to be carried out for small areas at which population data are available:

- In England and Wales, a Lower Super Output Area (LSOA) - or a combination of LSOAs
- Data zones in Scotland
- Electoral Divisions (EDs) for cancer incidence in Ireland
- Counties for cancer mortality in Ireland

Often we will want to include possible explanatory variables. Many of these are not available at the smallest area level and will have to be inferred from other data sources such as:

- smoking estimates at MSOA level
- alcohol/physical activity from the General Lifestyle Survey
- Health profiles/PHOF data at local authority level
- deprivation measures at LSOA level in England and Wales, at ED level in Ireland, and data zone level in Scotland.

Analysis of the data

Once the suspected cluster and reference areas have been defined, the first step in investigating a potential cancer cluster is the comparison of incidence and/or mortality rates in the two areas. It is known that cancer incidence varies by age and by sex, so it is important to account for differences in the age and sex distribution when doing this. This may be by direct comparison of age and sex specific cancer incidence rates, but initially it is convenient to calculate a single summary figure to allow easier comparisons to be made. From now on, only incidence rates are considered but the same methods apply equally well to mortality rates.

There are two usual approaches for accounting for differences in population structure between the two areas: indirect standardisation (SIRs/SMRs) and direct standardisation (ASRs).

Indirect standardisation produces results in terms of expected and observed numbers which is easier for a member of the general public to grasp. However areas are not directly comparable so they all relate back to one baseline (usually the national average which might not be very typical of the local area). Indirect standardisation produces stable estimates even when dealing with small numbers.

A consideration when using direct standardisation is that it is possible to convert the confidence intervals into a range of expected counts. That is, you can say: 'we would expect anywhere between 5 and 10 cases per year'. This is somewhat more intuitive to grasp.

Direct standardisation allows the comparison of several areas e.g. local, regional and national. This is often requested as part of analysis. However, it doesn't tell us how many people have been diagnosed in the area/population and the rate itself is rather abstract based on number of cases per 100,000 for example. Direct standardisation can also be unstable when dealing with small numbers.

Indirect standardisation

The indirectly standardised incidence ratio (SIR) is the ratio of the observed number of cancers to the expected number of cancers (APHO, 2008). The expected number of cancers is calculated by applying the age and sex-specific incidence rates of the reference population to the age and sex-specific population of the potential cluster. Table 1 shows a simple example; in a real investigation it is likely that more age groups would be used.

When calculating SIRs it can be useful to use a wider local area e.g. a city or region as one of the comparators. This helps to put the results into context when a whole area is different to the national average.

Table 1: Example calculation of a SIR

Sex	Age	Population	Reference Observed Cancers	Incidence (per 1)	Population	Cluster Observed Cancers	Expected Cancers
		Pr	Or	Or/Pr	Pc	Oc	(Or/Pr)*Pc
Male	**<50**	100,000	100	0.001	1,000	3	1
Male	**50-74**	50,000	200	0.004	500	3	2
Males	**75+**	25,000	300	0.012	500	2	6
Female	**<50**	100,000	150	0.002	1,000	2	2
Female	**50-74**	50,000	150	0.003	500	7	2
Female	**75+**	25,000	200	0.008	500	4	4
				Total		21	16
				SIR	21/16 =	131%	

The example in Table 1 shows that the SIR is higher than expected, but this may be due to chance. Confidence intervals for SIRs can be calculated either by Byar's approximation or an exact Poisson approach (APHO, 2008), discussed in the web-appendix with example approaches in Excel, Stata and R.

Direct standardisation

The directly standardised incidence rate (DSR) is the weighted mean of the age and sex-specific incidence rates. The standard weighting to use is the 2013 European Standard Population (Eurostat 2013), although in some cases the 1976 European Standard Population may still be used. In Table 2, the standard population weights are grouped to show a simple example.

Table 2: Example calculation of a DSR

Sex	Age	Standard Population	Population	Observed Cancers	Incidence (per 1)	Weighted Incidence	Population	Observed Cancers	Incidence (per 1)	Weighted Incidence
		S	Pr	Or	Ir = Or/Pr	Wr = S*Ir	Pc	Oc	Ic = Oc/Pc	Wc = S*Ic
Male	<50	61,000	100,000	100	0.001	61	1,000	3	0.003	183
Male	50-74	30,000	50,000	200	0.004	120	500	3	0.006	180
Male	75+	9,000	25,000	300	0.012	108	500	2	0.004	36
Female	<50	61,000	100,000	150	0.002	92	1,000	2	2	122
Female	50-74	30,000	50,000	150	0.003	90	500	7	2	420
Female	75+	9,000	25,000	200	0.008	72	500	4	4	72
	DSR				sum[Wr]/sum[S] =	0.003			sum[Wc]/sum[S] =	0.005

It is also possible to calculate confidence intervals for the directly standardised rate. The Association of Public Health Observatories (APHO) have recommended that Dobson, Kuulasmaa, Eberle and

Scherer (DKES) confidence intervals (Dobson, 1991) be used (APHO, 2008). However, Fay and Feuer (1997) proposed a set of confidence intervals based on the gamma distribution which were subsequently improved by Tiwari, Clegg and Zou (2006), and it is recommended that Tiwari, Clegg and Zou confidence intervals are used, as these are more reliable than DKES intervals in cases where the potential cluster population has substantially different age and sex distribution than the reference population. Methods of calculating Tiwari, Clegg and Zou confidence intervals in Stata and R are shown in the appendix. Excel is not recommended for use in calculating Tiwari, Clegg and Zou intervals.

Adjusting for factors other than age and sex

We can use either indirect or direct standardisation to adjust for factors other than age and sex. Often, we also adjust for deprivation, although the two methods discussed here are appropriate for other factors as well, so long as factors can be easily categorised.

A straightforward way of adjusting for deprivation would be only to compare the cluster area with areas in the rest of the country which are similarly deprived. This may be preferable if the area in the suspected cluster is all of a similarly deprived, because it is simpler for analysis and easier to describe what has been done.

Sometimes, areas in the suspected cluster cross have very different levels of deprivation. There is nothing special about standardising by age and sex; we may standardise by only age or only sex, or by additional variables such as deprivation. If we do have a suspected cluster covering areas of very different deprivation, it is straightforward to produce SIRs or DSRs standardised by age, sex and deprivation quintile by stratifying the data for the suspected cluster and the reference population by deprivation as well as age and sex.

Presenting the results

A map of the area investigated is valuable for public health and third parties alike to understand which areas are and are not covered by the statistical areas chosen for analysis. In many cases the latter have meaningless identifiers which do not tie in with a local sense of 'place' or well-known place names.

Care is required in writing up results as the report may be read by a lay audience or by public health professionals who are not familiar with statistical techniques. The principles of a 'Plain English' approach are likely to be helpful.

Further investigation

Further investigation will depend on the results of the analysis undertaken and the strength of evidence for a putative cause. This will generally require liason with the local authorities, those responsible for health protection, those responsible for environmental hazards, and - in England - the PHE Centre.

In the event of a positive result which is inconsistent with any data held on lifestyle (e.g. smoking rates) and/or has a number of cancers which are known to be linked to any suggested environmental or occupational cause (e.g mesothelioma and asbestos) then there needs to be a consensus decision on whether further detailed epidemiological investigation is undertaken. The public health professional in charge should coordinate a group (which does not necessarily need to meet face-to-face) representing local authorities, the cancer registration organisation, those responsible for health protection, and those responsible for environmental hazards. This group should review the evidence available, including the analysis undertaken already, to come to a decision.

Analysts working with cancer registration data cannot generally provide sophisticated spatial analytical skills, but can provide advice on what methods are available, the resources that can be utilised and the organisations that may be able to provide assistance with such analyses.

A protocol

Based on experience and a presentation delivered by the South West Public Health Observatory (SWPHO) in 2004, it is possible to define a brief protocol to follow with regards to a request:

- liase with the relevant authority responsible for local public health to establish who will be the main contact with enquirer, and decide if any investigation/analysis is warranted.
- where an environmental hazard has been identified as a cause of the cluster you should inform the local health protection team that you have an enquiry and give them the name of the local DPH who is acting as contact. Responsibility for health protection varies across the UK and Ireland:
  - o In England there are 12 Centres run by PHE
  - o The Public Health Wales Communicable Disease Surveillance Centre (CDSC), part of the Health Protection Team, is the epidemiological investigation arm of Public Health Wales. It protects the population from infection through surveillance of infectious disease, support for outbreak investigation, provision of health intelligence and applied research.
  - o Scotland: Health Protection Scotland at national level; NHS Boards at local level.
  - o Northern Ireland Public Health Agency
  - o HSE (Health Service Executive) in Ireland

- define the investigation (location, ages, sites, time period, geography). Are these decisions justifiable?
- has the request been raised as a result of media (mis)information? Get a copy of any press cuttings and be aware of what has been said in relation to the issue. If the issue appears to be misrepresented in any way an early press statement from public health team helps keep things calm while the data is being analysed.
- do an overview analysis of ASRs/SIRs. Are the results statistically significant? Are all cases recorded?
- a brief overview of potential carcinogens and biological plausibility. This would be an ideal point to highlight the main public health risk factors for cancer, and how people can get help to reduce them e.g the NHS Stop Smoking Service
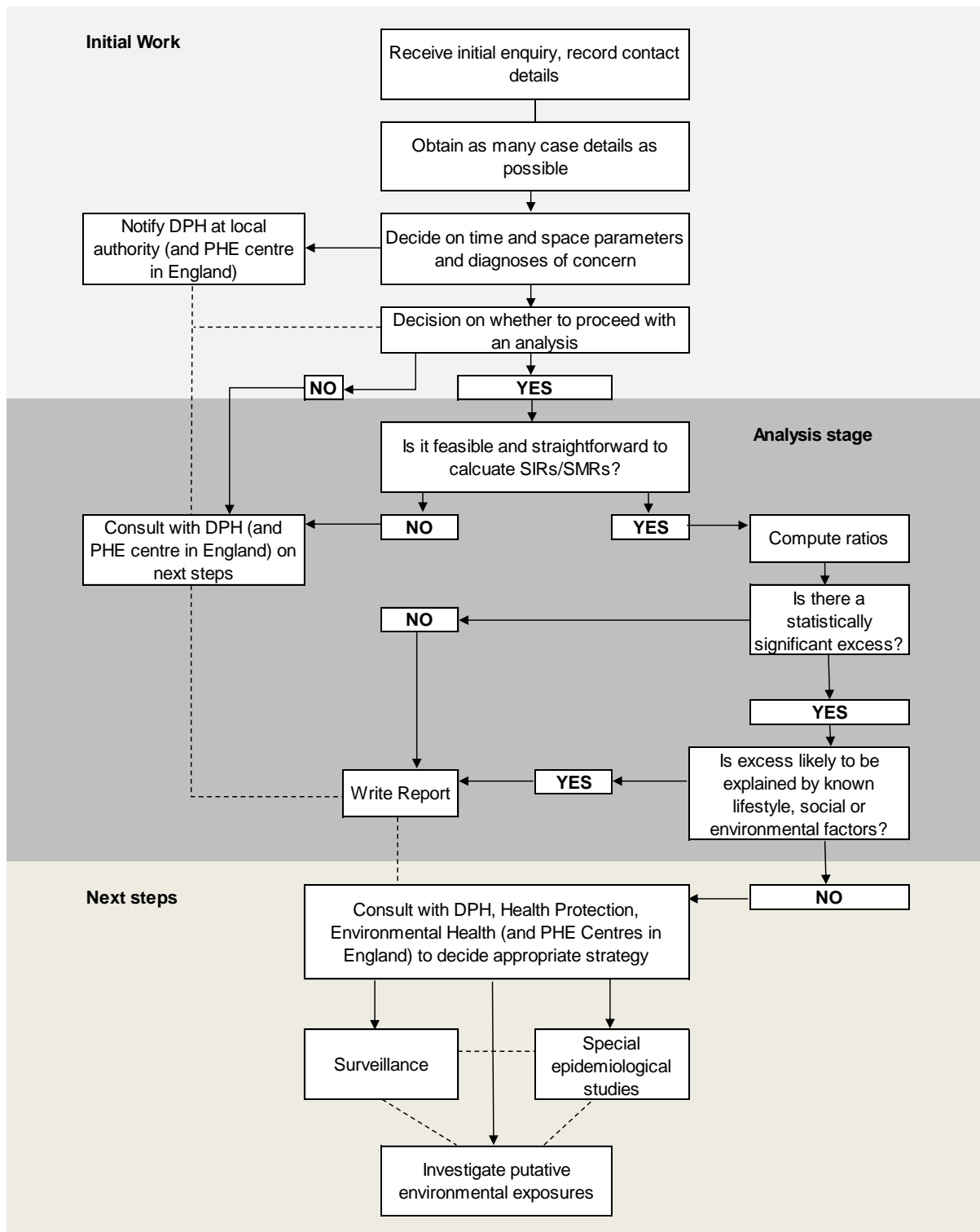
Figure 1 attempts to show the process in a simplified flowchart.

Example of a real request

The following request was received by PHE KIT South West in January 2013:
- residents voiced general concerns on pancreatic cancer in Langport, Somerset to a council official
- lack of public contact to define area, so chose two LSOAs to cover all town
- defined deaths based on ICD-10 code C25 from mortality database
- small numbers so combined 2002-2011 data (ten years) and compared to Somerset and South West
- noted strong smoking link and advised to contact GP for smoking cessation advice

Figure 1: Protocol flowchart for investigation of a cancer cluster using registration data

Initial Work

Receive initial enquiry, record contact details

↓

Obtain as many case details as possible

↓

Notify DPH at local authority (and PHE centre in England) ← Decide on time and space parameters and diagnoses of concern

↓

Decision on whether to proceed with an analysis

→ **NO**

↓ **YES**

Analysis stage

Is it feasible and straightforward to calculate SIRs/SMRs?

NO ← → **YES** → Compute ratios

Consult with DPH (and PHE centre in England) on next steps

↓

Is there a statistically significant excess?

NO ←

↓ **YES**

Is excess likely to be explained by known lifestyle, social or environmental factors?

Write Report ← **YES** ←

↓ **NO**

Next steps

Consult with DPH, Health Protection, Environmental Health (and PHE Centres in England) to decide appropriate strategy

↓

Surveillance — — — Special epidemiological studies

↓

Investigate putative environmental exposures

Other software

SatScan

It is sometimes desirable to add a more sophisticated level of analysis, perhaps if explicit objectivity is required. SatScan is one way of achieving this. This is software to objectively analyse number of cases, geographical proximity and population demographics. It outputs suspected clusters with a geographic centre, size and statistical significance. The output is mappable which helps with analysis.

The cases, time periods and choice of populations is user supplied, so there is still risk of bias if not selected appropriately.

The software is freely downloadable from: http://satscan.org

SAHSU RIF

The Small-Area Health Statistics Unit (SAHSU) at Imperial College London is due to launch their Rapid Enquiry Facility in 2015. This is an online tool which allows the analyst to enter their routine data and perform a more sophisticated analysis. This can help to confirm that a cancer incidence higher than expected does exist in the area under investigation.

Information is available at: http://www.sahsu.org/content/rapid-inquiry-facility

References

Association of Public Health Observatories. APHO Technical Briefing 3: Commonly used public health statistics and their confidence intervals. 1998. Available from: http://www.apho.org.uk/resource/item.aspx?RID=48457 [accessed 15th May 2015].

Armenian HK, Lilienfield AM. Incubation period of disease. *Epidemiologic Reviews* 1983.5(1):1-15.

Dobson AJ, Kuulasmaa K, Eberle E, Scherer J. Confidence intervals for weighted sums of poisson parameters. *Statistics in Medicine* 1991. 10(3):457-462

Fay MP, Feuer EJ. Confidence intervals for directly standardized rates: a method based on the gamma distribution. *Statistics in Medicine* 1997. 16:791-801

Parkin DM, Boyd L, Walker LC . The fraction of cancer attributable to lifestyle and environmental factors in the UK in 2010. *British Journal of Cancer* 2011. 105:S77-S81

Tiwari RC, Clegg LX, Zou Z. Efficient interval estimation for age-adjusted cancer rates. *Statistical Methods in Medical Research* 2006. 15:547-569

Standard text

This is a collection of text which can be reused in reports to offer a consistent message on our approach to these requests. Each paragraph may overlap with others so checking for duplication and consistency is required.

Small numbers and chance variation

• When looking at data in small areas there can be short-term fluctuations or apparent grouping of cases and deaths by chance. These can cause concern but are a normal outcome of complex natural processes (as many health issues are). Therefore, statistical comparisons have been made to test the likelihood of the results being due to chance variation or some other cause.

Age/sex standardisation

• The risk of developing or dying from cancer is strongly influenced by age and sex, so the make-up of the population in an area will affect the number of cancer cases and deaths observed. The incidence and death rates reported here have been standardised to the European Standard Population 2013 to take into account differences in age and sex.

• To allow a fair comparison between different areas, or the same area through time, directly age-standardised rates (ASRs) are calculated. This method looks at the occurrence of cancer/death in the population of interest, and calculates what the rate would be in a pre-determined population (known as the standard population), controlling for age. If the same standard population is used for all populations analysed, then the rates are directly comparable with each other to measure whether there are any differences of significance. Rates are standardised to the European Standard Population 2013.

Confidence intervals

• Confidence intervals are used to define the degree of variation in rates that can be considered normal. In this report, we used 95% confidence intervals. Generally speaking, when the confidence intervals of two age-standardised rates overlap, there is no statistically significant difference between them.

• When age standardised rates are reported, the 95% confidence intervals are also reported (Lower Confidence Interval: LCI, Upper Confidence Interval: UCI). The confidence interval is a statistical concept that defines the degree of variation in a rate which can be considered normal. Occurrence of, or death from, a disease depends on complicated relationships between a multitude of risk factors, so there are fluctuations over time or geography that do not necessarily reflect a real increase or decrease in a rate. As the population of interest becomes smaller, an increase or decrease of just one or two

cases/deaths has a larger effect on the rate. Hence, the smaller the population and the smaller the number of cases/deaths, the larger the confidence interval.

Lower Super Output Areas (LSOAs)

•   LSOAs are small geographical areas defined by the Office for National Statistics (ONS) that are designed for statistical purposes. Unlike parishes of wards they have an optimum population of about 1,500 people. Each LSOA can be identified spatially by a boundary or by its 'population weighted centroid' – a single point representing the centre of the population, these are used to identify those LSOAs within a given area of interest. LSOAs provide the smallest statistical building blocks for examining cancer incidence although routine data will rarely be published for them, but instead for groups of several LSOAs known as Middle Super Output Areas (MSOAs). Analysis is usually performed for several LSOAs as it is not feasible to examine cancer incidence for a smaller area. This is because the numbers are likely to be very small, which may lead to issues regarding identification of patients and make any statistical analysis meaningless.

Malignant/benign cancers

•   Generally when analysing cancer incidence only primary malignant tumours are counted, but this is not to say that benign tumours do not cause concern for patients or impact on the health system. Here we have analysed ....(complete as appropriate).

Cancer risks

•   On an individual level, the biggest risk of getting cancer is increasing age. Half of all cancers are diagnosed in those aged 70 years and older and three-quarters in those aged 60 and older. After age, the main cancer risks are: smoking (responsible for 20% of cancers); excessive alcohol use; being overweight; occupational exposure (primarily for men); and infections (primarily for women) (Parkin, 2011).

•   Those who are concerned about developing cancer should talk to their GP about ways in which they can modify their lifestyle to reduce their risk. The most important of these is stopping smoking. It is also important to attend cancer screening when invited, as these programmes have been evidenced to reduce cancer mortality.