Investigating
Clusters of
Non-Communicable Disease

Guidelines for Public Health Services
# Table of Contents

List of Figures  
v  
Acknowledgements  
i  
Preface  
vi  
Executive Summary  
vii  

## Background

- Introduction  
- What is a Cluster?  
- Examples of Non-Communicable Disease Clusters  
- Clusters in New Zealand  
- Surveillance for Clusters  
- Occupational Clusters  

## The Cluster Investigation Process

- Introduction  
- Stage 1 – Preliminary Evaluation of a Report of an Alleged Cluster  
- Stage 2 – Verification of Index Case and Exposure Reports  
- Stage 3 – Full Case Ascertainment  
- Stage 4 – Surveillance or Epidemiological Study  
- Legislation  

## Communicating Risk

- Introduction  
- Risk Communication  
- Perception of Risk  
- Communicating with the Public  
- Risk Comparison  
- Relationships with the Media  
- Conclusion  

## Appendix 1 – Cluster Report Form

## Appendix 2 – Data Recording

## Glossary

## References
List of Figures

Figure 1: Overview of the cluster investigation process
Figure 2: Stage 1 – Preliminary evaluation of a report of an alleged cluster
Figure 3: Stage 2 – Verification of index case and exposure reports
Figure 4: Stage 3 – Full case ascertainment
Acknowledgements

We wish to acknowledge the assistance of Associate Professor Neil Pearce, Wellington School of Medicine, and Dr John Harris, Director, California Birth Defects Monitoring Programme, for peer review.

Comments were also provided by Dr Martin Tobias and Henry Dowler, Ministry of Health.

We also wish to thank Dr Gillian Durham, Ministry of Health, and Medical Officers of Health who gave us feedback on a previous version of the guidelines produced by the Public Health Commission.

Deborah Reid MBChB FAFPHM FRACMA, Public Health Physician
Barry Borman PhD, Epidemiologist
Preface

The investigation of an apparent cluster of non-communicable disease, such as a cancer cluster or a cluster of birth defects, can be a complex and resource intensive task, and one that requires thorough planning and careful evaluation.

These guidelines are intended to assist public health services and in particular medical officers of health, to carry out efficient cluster investigations, address public concerns, and provide sensible advice.

A four stage investigation protocol is detailed, and a standard cluster investigation instrument is included. As each stage of the investigation is completed, a decision must be made whether to proceed to the next stage, with the final stage sometimes involving a formal epidemiological study. Criteria are suggested to assist in this decision making process.

Managing the concerns of the public and the media is an important part of cluster investigation. The affected community may feel outrage at the occurrence of an alleged cluster of disease, possibly attributed to exposure to a preventable hazard, and may demand action from the medical officer of health. The guidelines include suggestions for communicating risk, managing risk perceptions and feelings of outrage, and meeting the needs of the public and the media for information.

The guidelines are also available on the Ministry of Health’s Web site at http://www.moh.govt.nz

The Safety and Regulation Branch, Ministry of Health, would appreciate your comments of the implementation of the guidelines. Comments should be addressed to: The Director of Public Health and General Manager, Public Health Group, Ministry of Health, PO Box 5013, Wellington.

Dr Gillian Durham
Director of Public Health and
Deputy Director-General, Safety and Regulation Branch
The investigation of alleged clusters of non-communicable disease can be a complex and resource-intensive activity that requires thorough planning and careful implementation.

These guidelines provide a systematic approach to the investigation of clusters of non-communicable disease for public health agencies to carry out an organised and co-ordinated response to reports of alleged clusters.

Investigation involves four distinct stages. Each successive stage involves collecting more specific data and requires a greater verification of that data. However, a public health agency may choose to combine stages, depending on local judgement, experience, and the available resources. A decision to proceed with further investigation is made at the end of each stage.

Stage 1 gives the procedure to follow when an alleged cluster is initially reported to a public health agency. This involves recording the initial report, developing a case definition and follow up with the informant. Investigating a cluster suspected from monitoring or vital statistics begins at Stage 2.

In Stage 2 the index case(s) and exposure(s) are verified. Specific tasks to be carried out include deciding who should carry out the verification, literature review, and identification and review of the appropriate records.

Stage 3 identifies the confirmed cases in the time period and geographical area of the cluster and determines if there is a true cluster. A case finding team is formed, the case definition is revised, cases are identified and data collected, and the observed number of cases is compared with the number of cases expected in the time period and geographical area.

Stage 4 provides the option of continuing the investigation either by surveillance or an epidemiological study.

Dealing with the concerns of the public and media is fundamental to investigating clusters. The public often feel threatened about the occurrence of alleged clusters and demand action and information from the health professionals involved in an investigation. The guidelines include sections on risk communication, risk perception, and advice for handling inquiries from the public and the media.
Background

Introduction

The need to follow up reports of alleged clusters of disease can be expected to increase with increasing public awareness and concern about environmental exposures. This task often has to be done in the glare of publicity and under urgent and stressful circumstances.

Although Rothman (1990) maintains there is little scientific value in cluster investigations, others contend they should viewed on more than their scientific merit alone. Cluster investigations are an important public health strategy for responding to public concern about possible associations between disease and environmental exposures (Fiore et al, 1990; Neutra, 1990; California Department of Health Sciences (CDHS), 1989; Centers for Disease Control (CDC), 1990).

Public health agencies need to recognise the social dimensions of a cluster, how risks are perceived by the community, and the influence of the media on that perception. From a public health perspective, a community’s perception of a cluster may be more important than establishing the scientific existence of the cluster. The general public may not be satisfied with complex epidemiological or statistical arguments that deny the existence or importance of a cluster. Achieving rapport with a concerned community is critical to managing the situation.

An investigation of a cluster of non-communicable disease may assist in generating hypotheses that can be tested in another study population and time period. These guidelines are intended to assist investigators respond more effectively and in a more timely manner to local community concerns about environmental health issues. They also give a clear plan of action that can be outlined to the first person to report an alleged cluster.

What is a Cluster?

The term “cluster” has been used to describe an aggregation of some relatively uncommon disease or event (Last, 1988).
The initial characteristics of a cluster are:

- there is a definable health event
- there are usually at least two cases of the health event
- there is a perceived closeness of the cases within a time period and/or area defined by the informant
- a potential exposure is suspected, along with an alleged connection between the exposure and the health event
- the situation is generally unusual or unexpected
- the informant or the community requests some explanation of the health event.

Three categories of clusters may be reported:

- time clusters – when an unusual number of cases of a disease(s) occurs within a defined period of time
- space clusters – when an unusual number of cases of a disease(s) occurs within a defined area
- time-space clusters – when an unusual number of cases of a disease(s) occurs within a defined time period and area.

**Examples of Non-Communicable Disease Clusters**

Non-communicable disease cluster investigations reported in the scientific literature cover a broad range of health events, both acute and chronic. The Minnesota Department of Health received 420 reports of alleged clusters between 1981 and 1988. Most of these reports involved cancer (Bender et al, 1990). A statewide survey estimated an average of about four requests a day in the United States for cancer cluster investigations. The majority of inquiries ended with a response to the informant based on the information collected from the informant and in some cases examination of readily available data. Depending on the state, only one to three percent resulted in an epidemiological study or environmental sampling (Greenberg and Wartenberg, 1991).

Chance is the most frequent explanation of clusters (Neutra, 1990). An excess of cases was shown in only 16 of 61 cluster investigations in the United States, and no specific causal agents were identified (Schulte et al, 1987).
However, many carcinogens have been identified because of occupational or medical clusters. A literature search by Neutra (1990) found only one neighbourhood cancer cluster investigation that identified a carcinogen. This investigation discovered an association between exposure to the mineral erionite and mesothelioma in a Turkish village.

Evaluation of clusters has identified important causal relationships. Examples include birth defects and thalidomide, vaginal adenocarcinoma among young women who had in utero exposure to diethylnylstilboestrol, angiosarcoma of the liver and vinyl chloride exposure, male infertility and exposure to the pesticide dibromochloropropane, eosinophilia-myalgia syndrome and L-tryptophan (McBride, 1961; Lenz, 1962; Herbst et al, 1971; Creech and Johnson, 1974; Whorton et al, 1977; CDC, 1989). Fleming et al (1991) identified 87 cluster reports from 1775 to 1990 that established causal relationships in occupational medicine. Many other reported clusters have had no obvious common aetiology or have been shown on further investigation to have no more cases than would be expected in the general population.

Characteristics that increase the likelihood of a cluster investigation detecting a causal agent are:

- there are at least five cases and a very high relative risk (e.g., \( RR \geq 20 \))
- a unique and detectable class of agents has been responsible for the disease in the past
- the pathophysiological mechanism for the disease is well understood
- the agent persists in the environment and can be measured there
- the agent persists or leaves a physiological response in the individuals who are exposed and is rare in the normal population
- exposure is heterogeneous within the community
- accurate self-exposure assessment is possible by questionnaire or can be obtained from records
- a multi-community study consisting of some similarly exposed and unexposed communities is feasible
- the cluster represents an uninvestigated endemic space cluster rather than a space-time cluster. This scenario suggests a stable, persistent problem and possibly a persistent agent (Neutra, 1990).
Clusters in New Zealand

Few reports of clusters in New Zealand have been published other than in the media. Sare and Forbes (1972) reported a cluster of spina bifida from adjoining farms in the Waikato and suggested the herbicide, 2,4,5-T, as a causal agent. In 1977 the public identified further alleged clusters of spina bifida in Taranaki, Northland and the Waikato. A Department of Health inquiry found no excess incidence of spina bifida and no evidence to implicate 2,4,5-T as a causal agent (Anon, 1977). In 1990 the Department of Health investigated a reported cluster of congenital cataracts in the Wellington area. No excess incidence and no common aetiology was demonstrated (Elwood, 1990). A study of the health status of fire fighters involved in the ICI chemical fire found a cluster of testicular cancer among the comparison group of Wellington fire fighters (Bandaranayake et al, 1993). Subsequently another case of testicular cancer in a Wellington fire fighter was reported to the Fire Service. Investigation of the cluster showed there was an excess incidence of testicular cancer in the 1980s but no obvious common aetiology (Bates and Lane, 1995).

Clusters of suicide in prisons and police cells and of attempted suicide among young people have been identified in New Zealand from analysis of mortality and hospitalisation data (Cox and Skegg, 1993; Gould et al, 1994).

In 1993, an alleged cluster of birth defects in the children of three former Christchurch City Council horticultural workers received intense media scrutiny (Borman and Read, 1995). Allegations of an association with exposure to a fungicide, benomyl, led to the voluntary temporary removal of the product by some retail chains. An independent inquiry carried out for the Council found no evidence of an excess rate of birth defects and no evidence linking the birth defects to benomyl exposure (Alchin, 1994).

Surveillance for Clusters

Active surveillance to identify clusters is best carried out in the workplace where the population at risk and the exposures are limited and can be defined. In contrast, routine analysis of registries or vital statistics by public health agencies for unsuspected clusters is not recommended as a valid public health exercise (Smith and Neutra, 1993; Elliott et al, 1995). Reasons given by Smith and Neutra (1993) include:

- cluster reports from the public still require a response
- most diseases do not have a timely up-to-date registry
- information routinely collected by registries is often inadequate for investigating clusters (eg, address at the time of diagnosis may not be aetiologically relevant)
• analysis tends to focus on geographical clusters and may not identify occupational or other clusters

• routine analysis identifies false positives as well as false negatives. The cost of a false positive to a community may be considerable in terms of outrage, concern about personal health and loss of business or falling property values.

Smith and Neutra (1993) advocate vigilance to detect new and unusual exposures and evaluation of their impact as a more appropriate surveillance activity by public health agencies.

**Occupational Clusters**

Occupational clusters are likely to continue to contribute to the understanding of relationships between disease and exposure that are not predicted by toxicological and epidemiological studies at the time new chemicals and work processes are developed.

Advantages of the workplace with respect to cluster investigation include natural denominator boundaries as a result of shared work areas and identifiable time periods associated with work processes, shared exposure(s), the ability to form hypotheses based on job descriptions, and the possibility of finding comparable populations in which these hypotheses can be tested (Fleming et al, 1991).

Methods have been developed for the initial assessment of occupational cancer clusters using limited data such as basic data about the cluster and the size of the workforce, and the number of workers entering and leaving the workforce in each year (Smith et al, 1994). A step-by-step protocol for the investigation of disease clusters in the workplace has been developed by Fleming et al (1992).
The Cluster Investigation Process

Introduction

There are four distinct stages in the investigation of a cluster of non-communicable disease (Figure 1). Each successive stage involves collecting more specific, but a wider variety of, data and greater verification of that data. The boundaries between these stages are flexible. Depending on local judgement, experience, and the available resources, a public health agency may choose to combine stages.

At the end of each stage a decision must be made about whether to proceed further, and how to communicate the results of that decision to the public and other interested parties.

This step-by-step approach follows the principles of epidemiological research:

• establish that a problem exists
• confirm the homogeneity of the events
• collect data on all events
• characterise the events as to epidemiological factors
• look for patterns and trends
• formulate a hypothesis
• test the hypothesis
• write a report, obtain peer review and communicate the results.
Figure 1: Overview of the Cluster Investigation Process

Complaint report received → Collect basic information by phone and/or mail

Stage 1
- Further investigation needed?
  - No → Index cases and/or exposure verified?
    - No → STOP Write a report Communicate findings to informant
    - Yes → Verify index cases and exposure
  - Yes → Consult and review literature

Stage 2
- Index cases and/or exposure verified?
  - No → STOP Write a report Communicate findings to informant
  - Yes → Develop communication with community

Stage 3
- Full case ascertainment needed or feasible?
  - No → STOP Write a report Communicate findings to informant
  - Yes → Full case ascertainment
  - Yes → Obtain population and comparison data
  - Analysis
  - Consult and review literature

Stage 4
- Special study needed?
  - No → STOP Write a report Communicate findings to informant
  - Yes → Surveillance
  - Epidemiological study
Stage 1 - Preliminary Evaluation of a Report of an Alleged Cluster

The first stage gives the procedure to follow when an alleged cluster is initially reported to a public health agency (Figure 2). If a cluster is suspected from monitoring or vital statistics, the procedure begins at Stage 2.

**Step 1: Record the initial report**

Alleged clusters can be identified by anyone, including members of the public, news media, health professionals, local, regional or national agencies, environmental or health monitoring systems and vital statistics.

The public have a serious, quick and often adverse reaction to any thought of a cluster. They want the matter to be treated seriously and with concern by the public health authorities. Much of the success in managing public outrage depends on how the initial report is handled.

Whoever receives the report should identify himself or herself to the informant and tell the informant what actions will be taken, how long these will take, and when a response can be expected.

To avoid overlooking vital data, it is advisable to use a standardised form (Appendix 1) to collect this initial information. The informant’s real concern may also only emerge in response to careful questioning.

When an alleged cluster is initially reported, get as much information as possible about the informant and the index case(s). This can save time and resources later and also indicates that the report is being treated seriously.

**Step 2: Form an initial case definition**

The initial case definition is based on the following “what”, “where”, “when”, and “who” questions.

- What is the specific disease, symptom or health event of concern?
- Where is the affected geographical area, population group or workplace?
- When did the specific disease, symptom or health event occur?
- Who (eg. age, ethnicity, sex) are the index cases (ie, the cases first reported)?
- What, if any, are the suspected specific exposures?
Figure 2: Stage 1 - Preliminary Evaluation of a Report of an Alleged Cluster

STEP 1
Take initial report

STEP 2
Form an initial case definition which includes:
* what is the illness or symptoms?
* where is the affected area/population?
* who are the index cases?
* any suspected specific exposures?

STEP 3
Follow up with the informant

STEP 4
Consult and review information

Further investigation warranted?

CRITERIA FOR MAKING A DECISION
* unusually high number of cases
* a biologically plausible exposure
* intense community concern

STOP
* Write a report
* Communicate findings to informant

STAGE 2
Verify index case and exposure reports
Step 3: Follow up with the informant

Many of the reports of an alleged cluster can often be resolved at this step, either by an explanatory letter or telephone call. The investigation of clusters should have a strong health education component and community involvement (Neutra, 1990). Anxiety about a cluster can often be lessened when the informant is told that:

- a disease which the public perceives to be rare occurs quite often. For example, cancer is a relatively common disease and the risk increases with age; major birth defects occur in one to two percent of livebirths
- the length of time cases live in the cluster area must be substantial to implicate a plausible environmental carcinogen, because there is a long latency for most known carcinogens
- cases that occurred among people who are now deceased may not be helpful in linking exposure to disease because of the lack of data on exposure and on possible confounding factors
- the occurrence of clusters of specific diseases in a population is often due to chance.

There will rarely be a single explanation for the occurrence of a cluster. Many cluster investigations of non-communicable disease are initiated because one factor, often an easily identified environmental hazard, is suspected as the cause. However, most diseases have a number of possible causal factors, involving an interaction between genetic and environmental factors. For example, even though thalidomide is one of the most powerful teratogens, not all pregnant women who took it gave birth to an infant with birth defects.

Clusters with few cases are more readily explainable when the causal agent produces a five- to 10-fold excess rate (e.g., diethylstilboestrol). Most types of cancer occur at a rate of about one per 100 000 person-years.

Step 4: Consult and review information: make a decision

The decision to further investigate the alleged cluster is made after reviewing the initial information and consulting with specialists in cluster investigations and in the disease(s) concerned. Let the informant know the outcome of the decision as soon as possible.

In general, further investigation is warranted if there is:

- an unusually high number of cases
- a biologically plausible exposure(s); or
- intense community concern.
If it is decided that the alleged cluster demands further investigation, the informant could become an active participant in the data collection relating to the cluster. A member of the public may have easier access to some data that would be helpful to the investigation.

The following actions are recommended if a decision is made to end the investigation:

- write a report with a summary and conclusion
- obtain appropriate peer review
- communicate the results to the public.

This could involve writing a letter to the informant, or a public announcement, complete with press releases and public meetings. A written response to the person or organisation who initially reported the cluster will reduce the possibility of any misunderstanding about what was done and when.

Many reports of alleged clusters arise from genuine concerns of the public about either their current or future health. They want action, answers, and reassurance. All reports need to be treated seriously until there is evidence to the contrary. Public concerns can often be allayed by prompt action, keeping the public informed, and openly providing as much information as possible.

**Stage 2 - Verification of Index Case and Exposure Reports**

The next stage in the investigation is to verify the index case(s) and suspected exposure(s) that have been reported (Figure 3). In many instances the resources required to undertake an extensive verification are not available locally, and assistance from other specialists and agencies is needed.

**Step 1: Establish who should do the verification**

To verify the index case(s) it is advisable to consult with appropriate specialists (eg, a pathologist, neurologist, toxicologist) or specific agencies (eg, Occupational Safety and Health, Ministry of Health). An environmental health specialist or an occupational hygienist may be needed to verify the possibility of exposure. Often a perceived exposure is not verifiable (Fiore et al, 1990). Measurements of exposure are usually not necessary at this point.
Figure 3: Stage 2 - Verification of Index Case and Exposure Reports

Verification of index case and exposure reports

**STEP 1**
Establish who should do the case verification

**STEP 2**
Review the literature

**STEP 3**
Identify records that can be used for verification

- Establish measures to protect confidentiality of cases
- Develop communication with the community

**STEP 4**
Review the appropriate records and verify cases and exposure

Summarise

**STEP 5**
Cases and/or exposures verified?

- Yes
  - Consult
  - **STAGE 3** Full case ascertainment
- No
  - STOP
    - Write a report
    - Communicate findings to informant

**CRITERIA FOR MAKING A DECISION**
- index cases cannot be verified (informant had wrong diagnosis or illness began before suspected exposure)
- there is clearly no possible source of exposure
Step 2: Review the literature

The investigator also reviews the literature for evidence of any previously reported clusters of the disease, known exposure associations and other epidemiological and toxicological information. CD-ROM databases such as HAZARDTEXT or TOMES produced by MICROMEDEX may be useful when exposure to specific hazardous substances has occurred or is suspected.

Step 3: Identify the records that can be used for verification

Records that can be used to verify cases are:

- death certificates
- birth certificates
- hospital discharge records and case notes
- population-based cancer and birth defect registries
- union or company employment records
- doctors’ records.

Exposure is relatively easy to determine in acute disease clusters and can be done by questionnaire. In most instances, exposure has been through personal contact or through food, drug or beverage consumption. In contrast, exposure through water, air, soil or dust is poorly correlated with questionnaire responses (Neutra, 1990).

Potential exposure(s) can be verified from agency and company files about sites and facilities in the area of the cluster or where the index cases worked or lived, aerial photographs, records of water, soil and air quality from various monitoring agencies, and planning records about previous industrial sites and property uses.

Availability and access to information may be constrained by the Privacy Act 1993 (see Legislation). The informed consent of the index case(s) (or their next of kin if they are dead), may be necessary to access records.

Step 4: Review the appropriate records and verify cases and exposure

The easiest way to verify a disease is often by reviewing hospital records, in particular any diagnoses on pathology reports. Diseases can also be verified from doctors’ records and disease registries. Cause of death can be confirmed using the death certificate and reviewing appropriate medical records. The case definition provides guidance in deciding what data are to be used.
Verification of exposure is often more difficult because of the comparative lack of relevant data. It may be helpful to consult with epidemiologists and occupational and environmental health experts for information about the availability, access to, and use of data such as union or company employment records, and residential histories.

Early in the investigation of a cluster, there may be requests for new environmental data to be collected. Premature environmental measurements should be avoided, since they may be unfocused and uninterpretable.

**Limitations of records**

Much of the available data will have been collected for purposes other than for investigating a cluster. As a result, the recording of information may vary between sources and make the information difficult to interpret. For example, the hospital discharge form’s diagnosis might be pneumonia with no indication that the condition may be the result of exposure to a toxic substance. An individual’s cause of death also may not be identical to the reason for hospitalisation (eg, a subsequent pathology report could indicate a previously unreported tumour).

Many studies have shown that the level of disease ascertainment is directly related to the number of records used in searching for cases. Studies using multiple sources have a higher validity and level of ascertainment than studies involving only a limited number of sources.

**Step 5: Make a decision: stop and report or investigate further?**

At this stage of the investigation it is helpful to summarise in writing and review the findings to date. This may stimulate new ideas about the disease or possible causes.

Suicide clusters differ from other clusters, because the community’s perception that a cluster exists may itself be an important risk factor for further suicides and attempted suicides. Action is required, regardless of the number of cases reported. A community response should be initiated to identify and refer individuals at high risk of suicide for assistance (O’Carroll and Mercy, 1990).

Reported clusters will fall into three categories:

1. **No apparent cluster**

   The initial investigation will often find no apparent excess number of cases. In many instances the original disease and/or exposure allegations are not supported by the medical records and/or environmental inspection, or else the examination of records does not confirm the suspected environmental exposure.

   The possible effects of migration are also important. Some of the cases involved in the reported cluster may have developed the disease before moving into the area and encountering the possible exposure, and they should not have been included.
The disease or exposures alleged in the reported cluster may be a number of different diseases or exposures. The term “birth defects” includes a wide range of specific defects which have different epidemiological characteristics and are likely to have different aetiologies. When the diagnoses of the reported cancer cases have been verified they may be different types of cancer or not cancer. It is unlikely that unrelated cancers will constitute a cluster.

2 Explained apparent cluster

Many reports about clusters of cancer, spontaneous abortion, and birth defects result from the public not realising how common these conditions are. For example, after a clear explanation, the public are likely to understand that a few cases of lung cancer in a retirement community with a high percentage of smokers and no unusual environmental exposure is not likely to constitute a cluster. A high number of Down syndrome births in a local population with a high proportion of older mothers is also not uncommon.

The investigation can be stopped if either of the above options are found. A written report of the investigation to date and justification of the decision to stop should be made.

3 Unexplained apparent cluster

If the reported disease and/or exposure are confirmed, the investigator must decide to proceed to Stage 3 (full case ascertainment) or not. This decision depends on the type of disease(s) and exposure(s), the size of the apparent cluster, and the biological plausibility of a disease/exposure relationship.

To avoid the possibility of misinformation or confusion it is preferable to let the original informant know the decision by both:

- a personal telephone call, which gives them the chance to ask questions
- a following letter, to document the information clearly.

Stage 3 - Full Case Ascertainment

This stage involves finding and verifying all additional unreported cases of the disease(s) in question in the time period and geographical area of interest (Figure 4).

Step 1: Establish a case-finding team

It is often helpful to form a case-finding team because of the complexity of the environmental and occupational health issues and the need for a variety of disciplines, perspectives, and skills to carry out and analyse the data. The team needs to include
specialists in epidemiology, environmental health (if exposure verification is needed), and public health at least. The roles and responsibilities of team members, the channels for communication between members, and the spokesperson to the press or public should be decided at the outset.

**Step 2: Revise the case definition**

It is crucial at this stage of the investigation to review and if necessary revise the initial case definition. A complete case definition includes:

1. a definition of the health events to be counted
2. a time period during which diagnosis occurred
3. a geographical area and/or population group of interest.

**1. Health events**

- Only the specific disease or closely related diseases suspected of clustering are counted. For example, if squamous cell carcinoma of the lung is reported, all types of primary lung cancer would be counted because they might be caused by the same type of exposure. Other types of tumours would not be counted. For birth defects, both livebirths and stillbirths should be included and, if possible, other types of reproductive outcomes (e.g., spontaneous abortions). If the investigation initially focused on spina bifida, all types of neural tube defects should be included. Many birth defects (e.g., anencephalus) have a higher rate in stillbirths than in livebirths. Failure to collect stillbirth data will result in the investigation excluding an important number of cases. It is also important to decide on the follow up period in clusters of birth defects. A number of birth defects (e.g., heart defects) are not diagnosed until after the first year of life, and there may have to be an extended follow up period.
Figure 4: Stage 3 - Full Case Ascertainment

Full case ascertainment

- **STEP 1**
  - Establish a case-finding team

- **STEP 2**
  - * Revise the case definition
  - * Define health events, time period, and geographical area

- **STEP 3**
  - Develop and implement the case-finding strategy:
    - * phase 1 - what records to be examined?
    - * phase 2 - what data to be collected?
    - * phase 3 - ethical approval?
    - * phase 4 - collect data

- **STEP 4**
  - Count and analyse case data
  - Consult and review literature

**STOP**
- * Write a report
- * Communicate findings to informant

**Further investigation warranted?**

- **Yes**
  - STAGE 4 Surveillance
  - STAGE 4 Epidemiology study

- **No**

**CRITERIA FOR MAKING A DECISION**

- * is the excess of concern?
- * is the excess worthy of further study?
- * is the exposure biologically plausible?
- * is the community still concerned?
- * is there a sudden increase or an increase over time?
- * area cases concentrated around suspected environmental hazards?
• A broader set of conditions should be counted if there is concern about general increases in diseases that might have resulted from exposure to a mixture of toxic chemicals, or if a number of unrelated diseases were reported.

• Cases which cannot be confirmed using medical records should be tabulated separately or not be counted.

2 Time period

• The best time reference point for a chronic disease is the calendar year of diagnosis, but occasionally only the date of death is known. A time period of possible exposure needs to be defined, and all cases diagnosed during that period need to be identified. Clusters of cancer also need to consider an appropriate latency period.

• When investigating clusters of birth defects, focus on possible exposures at or before the time of conception, rather than at the time of birth. Exposures at the time of birth will not necessarily be the same or at the same level as at the time of conception. Most birth defects have a critical period when the defect can occur. For example, neural tube defects occur within the first 28 days after conception, so the occurrence of these defects will not be due to any potential exposure after that time.

3 Geographical area or population group

• The community of interest (eg, suburb, city, health district or territorial local authority (TLA)) is the basic areal unit for data collection and analysis involving possible community exposure(s). It is often more meaningful to disaggregate heavily populated areas into subareas. If the focus is on the occurrence of disease among a specific population subgroup (eg, a particular neighbourhood or all women over the age of 35 years) the smallest unit that includes the entire group and for which statistical data are available should be used as the denominator. This unit is also used for collecting information about membership in the subgroup. Alleged clusters are unlikely to respect the administrative boundaries that have determined reporting of population data or health statistics (Elliott, 1995).

• If the focus of the investigation is potential exposure at the workplace, the work site itself is usually the basic unit of analysis. Specific occupations or subgroups working in high exposure areas may be defined, but counting cases in the entire workplace provides the basis for comparisons between subgroups.

• Whatever data are collected for the cases should also be collected for the comparison or unexposed population. The coding of data should also be consistent within and between these two groups.
• Domicile data should relate to the place of usual residence. This may not be the domicile at the time of diagnosis. Data on the length of residence in a particular area are also important, especially for diseases such as cancers, in which exposure may have occurred 10 or more years previously. Without these additional data a person who has lived one year in an area is assumed to have a similar degree of exposure to a person who has lived in the same area for 10 or 20 years. In reality, the person with a short residency in an area, may have been exposed to a possible environmental hazard in another area, or have not been exposed to an environmental hazard in the current area.

• In the New Zealand mortality statistics, domicile is given as the usual place of residence. There is no definition of what constitutes “usual”, so this can range from months to years or life. The address of a family member may also be given as the usual place of residence even though the case has only temporarily stayed there.

• In clusters of birth defects it is important to know the domicile of the mother at the time of conception, and if possible, during the year before the birth. For childhood cancers, attempt to establish possible exposures from the time of gestation. In occupational clusters, the work history, not only the most recent occupation, is relevant. Occupation is recorded in the national mortality statistics as the occupation at the time of death.

For example, in a cluster involving a pesticide-induced food illness, the case definition might include all individuals who experienced vomiting or diarrhoea within two hours of eating produce anywhere in a defined TLA. The initial case definition in a possible cancer cluster might include all cancers that occurred during the last five years in children aged less than 15, living in that geographical area before they were diagnosed.

**Step 3: Develop and implement the case-finding**

There are four phases in the process of finding the cases and collecting the data:

**Phase 1: Decide what records need to be examined**

All cases with the disease(s) that were diagnosed in the area or workplace and time period need to be identified. This usually involves finding and reviewing data from several sources, including:

• hospital logs and hospital discharge records
• death records
• population-based registries (eg, for cancer)
• hospital-based tumour registries
• union or company employment records
• birth certificates
• records of specialists and researchers in the field.

Case finding could also involve laboratories, pharmacies, radiologists, and direct appeals to general practitioners and the public.

Each data source has its own particular strengths and weaknesses. For example, hospital records are a good source of information, but may include unconfirmed cases, such as suspected diagnoses, and omit cases not diagnosed in the hospital or diagnosed in another hospital. Death records may have a vague non-specific diagnosis or may omit a diagnosis when it was not the underlying cause of death.

Disease registries are a good source of cases for full case ascertainment. However, they may not be available in the area for the time period of interest. In the absence of a registry, full case ascertainment is more difficult and resource intensive.

**Phase 2: Determine what data will be collected**

Data availability depends on whether the source of information is registries, hospital or medical records, or interviews. In general, the following is the minimum data that is collected for each case:

• name or other identifier
• date of birth
• ethnicity
• sex
• age at time of diagnosis and/or death
• residence at the time of diagnosis
• diagnosis and basis of diagnosis.

If possible, also collect data on:

• family history of the disease(s) in question
• known exposures, such as smoking
• length of residence at the current address
• history of past residence in the area of interest

• other relevant exposure variables – for example, playing in contaminated fields, drinking contaminated water, eating contaminated foods, overseas travel.

If investigating a possible exposure in the workplace, take a detailed occupational history. This includes:

• occupation of the case or parent

• employment history – job type, classifications, and duration at each position – as far back as possible.

Further workplace information such as the following is desirable:

• known work exposures to toxic substances – for example, asbestos exposure of the case or parent

• identification of potential hazards at the work site.

Phase 3: Obtain ethical approval, if required, to carry out the data collection

Phase 4: Collect the data

The data collection methods chosen depend on the type of data needed to count all suspected cases. Much of the basic information can usually be obtained by reviewing existing data. It is important to indicate which source document is used to obtain the data so that data validity can be assessed. Data must be recorded and coded in a systematic and consistent manner (Appendix 2).

A questionnaire should be designed so that all the necessary data are obtained in a clear and unambiguous manner from one interview (Appendix 2). Consultation with experts about the technical aspects of questionnaire design, pretesting, training of interviewers, and data coding and processing is advisable.

It is important during data collection that confidentiality is adhered to.

Step 4: Count and analyse case data

Once every effort has been made to find all the cases in the cluster population that conform to the agreed case definition, the data are examined and any duplicate case reports are eliminated.
Numerator data for calculating cluster and background disease rates can be obtained from the national health statistics – for example, mortality, hospitalisations, the National Cancer Registry, the mental health register, and special studies. Data specification (eg, by age groups and ethnicity) should be in the same format as for the denominator data. Rates are not routinely calculated for small areas such as a neighbourhood because the number of observed cases is usually too small for stable rates or meaningful analysis (Fiore et al, 1990).

Population data are necessary to calculate expected numbers of a disease based on published reference rates of disease. The expected number of cases can be compared to the actual number of cases observed in the study population to determine whether the community has experienced an excess rate of the disease.

Cluster investigations usually require detailed population data (eg, by sex, ethnicity, and age groups) for very small geographical areas. Such data can be obtained every five years from national census data. Data for non-census years can be requested from Statistics New Zealand. Using census data in a non-census year assumes that there has been no change in the demographic composition of the population. Denominator data can also be obtained from a number of government departments, including the Ministry of Education and Department of Social Welfare, and from TLAs.

In epidemiological terms, the number of cases in the study or cluster population is the number of observed cases. The number of expected cases of a disease is determined by multiplying a background or comparison rate of disease by the study population in the time period and geographical area that was used in counting the observed cases.

Once a complete (or virtually complete) count of cases has been established, it must be decided whether the number of cases that has been observed is different from that expected. Usually cluster investigations are concerned with determining if there is an excess number of cases in a local population. A finding of fewer than expected cases is, however, reassuring for the community.

Often a cluster is thought to be “real” if there is a statistically significant excess of cases. The statistical tests used in the analysis calculate the probability that the disease rates observed in the cluster would occur by chance alone. This will usually involve comparing the observed rate in a known population with expected rates derived from larger population surveys or disease registries.

As most diseases are not evenly distributed throughout a population, an observed excess (or deficit) of cases may, therefore, occur at random and the cluster not be aetologically important. There are many census areas and small towns in New Zealand, and hundreds of thousands of workplaces and social groups like clubs, and churches. All of these groupings are at risk of excess rates of a non-communicable disease, even if the distribution of the disease itself is random. The number of observed cases will rarely equal the number expected, even if there were no environmental cause. The pivotal question is “How far away from the expected number must the observed number be to make it a very unusual occurrence?”

Standard statistical and epidemiological techniques for assessing excess risk can often be used to evaluate reported clusters, but statistical significance should not be used as the sole criterion for investigating a disease cluster. A small observed number of cases
may be worth investigating if there is a biologically credible exposure present or there is intense public concern.

A useful first step is often to produce frequency tables of the disease and examine the related descriptive statistics. Mapping the data is also helpful.

Diseases will occur at different rates in different age, ethnic and sex groups. The calculation of expected numbers should take into account the possible effect of these possibly confounding factors on the occurrence of the disease. If there are sufficient cases and population this can often be achieved using some form of standardisation, direct or indirect.

If the number of health events is too small to show meaningful rates, pooling across geographical areas or time may be possible. Other commonly used statistical approaches include chi-square tests of observed versus expected frequencies based on the Poisson distribution for low frequency data and Poisson regression used for comparison of rates. Confidence intervals may be calculated for point estimates.

Evaluating a spatial cluster can be done by comparing the rate in the study area with that in adjacent census areas or changing the geographic scale at which the analysis is carried out – for example, health district, TLA, region, New Zealand. If a temporal cluster is being assessed, the occurrence in that time period can be evaluated in the context of previous or subsequent periods. When comparisons are made, the comparison population must be carefully chosen to ensure it has similar demographic or exposure characteristics.

Analysis of cancer incidence data at a range of geographical scales gives information that can deal with public concerns, prevent expensive and unwarranted epidemiological studies driven by public and political pressure and target appropriate cases for further investigation. For example, although New Jersey was found to have an excess of childhood and young adult cancers when compared to other states, further examination showed that the clusters were not related to degree of urbanisation or county boundaries, and were made up of very few cases and in very few municipalities (Schneider et al, 1993).

Statistical verification of alleged clustering may be very difficult. For rare events the time frame selected from which to determine the expected incidence will influence whether the cluster is found to be statistically significant. Short time frames may be misleading. Definition of the geographical boundaries may have a similar effect.

Problems may arise from statistical techniques used to detect clusters (Rothman, 1990; Wartenberg and Greenberg, 1992). It is often difficult to distinguish between events of epidemiological and public health importance and those that occur as a result of chance. Some techniques may not be sensitive enough to detect true aggregations, while others may detect aggregations whose epidemiological and biological significance is difficult to interpret. A review by Wartenberg and Greenberg (1992) of results for four commonly used methods under two alternative hypotheses of environmental exposure showed that false negative rates depended highly on the method used and the nature of the exposure pattern that was sought true clusters were detected only if enough cases had been observed and if the method most sensitive to the type of exposure pattern had been used.
The needs of public health authorities are not well met by the variety of statistical techniques available, and it is desirable to obtain statistical advice. Many alleged clusters require only basic data analysis. For rare diseases in small areas the alleged cluster may only be one or two cases and may disappear once case histories and diagnoses are verified, and recent migrants and other anomalies detected as each case is reviewed. Clusters may also disappear or reappear by changing the time, space, or time and space boundaries, by over- or underenumeration of the population at risk, or by choosing different sets of standard rates. Decisions are often implicit rather than explicit, as they depend on existing data (Elliott et al, 1995). Results should be treated cautiously until more about the sensitivity and specificity of the methods used is known (Elliott, 1995).

Brief descriptions and critiques of some of the available statistical techniques are given in the CDC’s Guidelines for Investigating Clusters of Health Events (CDC, 1990). Many of these techniques are included in the computer software package STAT! produced by Biomedware.

**Step 5: Make a decision: stop and report or investigate further?**

The decision to stop and report on the investigation or continue to the next stage depends on a number of factors. Further investigation is usually not required if:

- there is no excess disease, and no exposure, and therefore no biological plausibility
- there is no excess disease, a possible exposure, but no biological plausibility that the exposure could result in an excess
- there is excess disease, no identified exposure, and no biological plausibility that the excess rate results from an environmental exposure.

If there is an excess of cases:

- is it of concern?
- does it warrant further study?
- is the exposure biologically plausible?
- is there a sudden increase in cases in a recent period or have the cases been increasing over time?
- are cases more concentrated around suspected environmental hazards or in suspected occupational groups?

In general, a “yes” answer to these questions increases the need for further follow up (Stage 4).

If the decision is made to end the investigation:
Stage 4 – Surveillance or Epidemiological Study

There are two options if the cluster warrants further investigation.

1 **Surveillance**

Surveillance is a more appropriate approach than an epidemiological study where an excess number of cases is found in the cluster but where the excess is of low statistical significance or the exposure has weak biological plausibility. A surveillance programme run over several years determines whether cases are increasing over time and what their geographical distribution is. For example, monitoring of childhood rhabdomyosarcoma incidence in a North Carolina county was proposed for a five-year period after a cluster was confirmed and preliminary investigation had revealed no plausible environmental exposures (Grimson et al, 1992).

If there are no registry or vital statistics data available, a reporting system may have to be established to receive reports about the disease from the public or health professionals.

2 **An epidemiological study**

If there is an excess of cases and there is a biologically plausible connection between the cases and some environmental exposure, further investigation of the cases and their environment is warranted.

Further investigation may involve a case-control, cohort or cross-sectional study and can range from a few days of work to years and hundreds to thousands of dollars. Consultation with appropriate specialists and agencies is recommended. These include the Occupational and Environmental Health Research Centre, the four university departments of public health, the Institute of Environmental Science and Research, and the Ministry of Health.

**Legislation**
There is a range of legislation that may apply during a cluster investigation. These statutes include:

- Health Act 1956, including s29 (nuisances), 60-63 (water pollution) and 128 (powers of entry)
- Cancer Registry Act 1993
- Toxic Substances Act 1979
- Resource Management Act 1991
- Health and Safety in Employment Act 1992
- Hazardous Substances and New Organisms Act 1996
- Accident Rehabilitation and Compensation Insurance Corporation Act 1992
- Local Government Act 1974
- Privacy Act 1993 (and Health Information Privacy Code 1994).

The Privacy Act and the Health Information Privacy Code are subject to other countervailing legislation. If designated officers are acting according to specific statutory powers, the Act and Code are overridden. Care has to be taken that the actions remain within the scope of those powers; otherwise, privacy legislation must be considered. If difficulties or concerns arise, consult your local privacy officer and/or legal adviser.

The Privacy Act establishes the set of information privacy principles which govern the collection, holding, use and disclosure of information.

The Health Information Privacy Code modifies the privacy principles to deal specifically with issues that arise in the health sector. It applies to health information related to identifiable individuals and not to anonymous or aggregated statistical information that does not allow individual identification.

The Code contains 12 rules that apply to health information and health agencies. Rules 1-4, 5, 10 and 11, relating to collection, storage, use and disclosure, are the most relevant to cluster investigations.

There are a number of exceptions to the rules. Many of the rules are circumvented by the exception that states that if the information is to be used for research or statistical purposes and that it will not be published in an identifiable form, then the particular rule does not apply.

Exceptions to Rules 10 and 11 allow disclosure of information if the health agency holding information obtained in connection with one purpose believes on reasonable grounds:
• that its use or disclosure for any other purpose is necessary to prevent or lessen a serious and imminent threat to public health or safety, or

• that non-compliance with the rule is necessary to avoid prejudice to the maintenance of the law by any public sector agency.

When collecting information, it may not be practicable to contact the individual for consent or to get the information directly from them. The agency or informant providing the information must believe on reasonable grounds that one of the exceptions to Rule 10 or 11 applies or that a designated officer is collecting it according to statutory functions.

Ethical approval may be required if the investigation proceeds beyond Stage 2. If there is doubt about the need for ethical approval, consult the chairperson of the local ethics committee.
Communicating Risk

Introduction

As already noted, public health agencies are often alerted to the occurrence of a cluster by the public and media. Responding to the community concerns and fears aroused by a cluster is integral to a cluster investigation. Early, consistent, honest and open communication can help to assure that any anxieties the public may have are addressed during the investigation and that inconclusive results do not come as an unwelcome surprise. Information about progress with and the outcome of the investigation is vital.

Investigators need to be able to recognise the source of any community suspicions of deliberate delay and cover-ups, as well as of demands for unrealistic schedules and allocation of resources (Rothenberg et al, 1990).

Risk Communication

Risk communication involves an exchange of information between a public health agency and the public about the nature, size, and significance of a cluster, and if necessary the control of any health risk (Covello, 1995). The goal of risk communication is to establish trust and credibility. It is not a one-way process, informing the public of the scientific aspects of a cluster. The early stages of cluster investigations often depend on information obtained from the public and media, to build the file about the cluster.

The success of a cluster investigation does not depend on proving the cause of a cluster, but rather on reaching an outcome which is satisfactory for all groups involved. This involves the mutual understanding and resolution of any conflict between the public’s expectation of a cluster investigation and the science of carrying out the analysis within the limits of available knowledge (National Research Council, 1989). Successful risk communication achieves this.
Perception of Risk

In a public health context, risk is a combination of scientifically defined hazard and public outrage (Sandman, 1991b). Therefore, the public and scientists usually have different estimates of risk (Slovik, 1987). People are frightened and angered by risks which are not necessarily the same as those that kill them (Cohn, 1989). The public often feels threatened, powerless and outraged about a newly disclosed potential danger to their personal health, the public health and the environment (Rowan, 1996).

A number of factors characteristic of clusters of non-communicable disease increase public concern about risk. These include fatalities and injuries that are grouped in time and space, considerable media attention, risks that are unfamiliar and uncontrollable, risks that involve an involuntary exposure and delayed effects, and situations where children are specifically at risk (Covello, 1995). For example, in McFarland, California, where 13 cases of childhood cancer have occurred over 15 years in a town of only a few thousand, an account of the cluster has been dramatised for television and formed the basis of a musical play, and the town has been visited by a presidential candidate. No common cause has been shown (Smith and Neutra, 1993).

The most powerful factors influencing the public perception of risk are the trust and credibility attributed to the source of risk information. This in turn is based on perceived caring and empathy, competence and expertise, honesty and openness, and dedication and commitment of the source (Covello, 1995). In contrast, one of the least powerful factors is the scientific data about risk, although this has usually been the major component of communications about risks and clusters.

Communicating with the Public

Scientific reports, either oral or written, can allay community fears about a cluster, but can also create confusion, dissatisfaction and a clamour for continuing investigation. For example, a cluster investigation of spontaneous abortions found that risk perception and risk communication had an equivalent, if not, greater standing as did the epidemiology and exposure assessment components of the study (McDiarmid et al, 1994).

The public typically want simple, easy-to-read and comprehensible information from a credible source. Scientists may consider such messages as incomplete, inaccurate or even biased (Glanz and Yang, 1996). Conversely, scientists frequently deliver detailed technical statements that may be accurate and unbiased but are often complex and filled with jargon and uncertainty (Goldstein et al, 1992).
While quantitative probabilities are intrinsic to cluster investigations, the public usually communicate in qualitative expressions. People often find it difficult to understand the technical meaning of terms such as “unlikely”, “not statistically significant”, or “a probability of 0.05”. One study found that people were equally likely to interpret a “70 percent chance of rain” as “rain 70 percent of the time”, “rain over 70 percent of the area” and “70 percent chance of some measurable rain” (National Research Council, 1989).

Risk Comparison

Scientists frequently try to put a particular risk into perspective by comparing it with other risks. The major difficulty is to find risks that are sufficiently similar to make the comparison meaningful. Many of the comparisons are rejected by the public because they see the subjects chosen for comparison as being controllable by the individual (eg, motor vehicle crashes), optional (eg, smoking), remunerative to the individual (eg, work), or an act of nature over which no individual or agency has control (eg, earthquakes) (Goldstein et al, 1992). It is preferable to compare risks of similar size that occur in the same context or have a similar outcome (National Research Council, 1989). The public may also view risk comparison to be more strategic than scientific, and more like propaganda than a public service.

Relationships with the Media

Investigators need to remember that the media amplify public outrage but do not create it (Sandman, 1991b). Journalists will usually want a story to have a visual component and to contain blame, politics, controversy and a strong emotive content. The media need to simplify complex and technical explanations and tend not to include scientific qualifications or subtle distinctions. Newspaper coverage of clusters in the United States rarely includes mention of multiple adverse health outcomes, confounding variables, relative risk, and data reliability. Scientists, on the other hand, usually believe accurate information about risks should be the most important feature of a good news story and that more epidemiological and risk information should be included (Greenberg and Wartenberg, 1990).

The media’s role may change with the size of the community it services. Newspapers in large metropolitan areas of the United States were more likely than newspapers in smaller communities to link contamination from local agents to threats to human health in the community and frame the story as problems. However, small community newspapers were more likely to frame local contamination in the context of solutions to the problem and to link contamination to health risks if the contamination was in a distant community (Griffen et al, 1995).
Source credibility is also a major factor in developing a successful relationship with the media. If possible get the support of environmental groups early in the investigation. For example, McCallum et al (1991) found that environmental groups had consistently high rankings for perceived expertise and credibility. Officials in the chemical industry were the most knowledgeable concerning the risks of chemicals in the community but were the least trusted by the public.

Conclusion

Oral or written reports of cluster investigations need to be presented in a manner that will not lead to confusion or distortion by the public or media. The messages should be clear and unambiguous, use plain language, avoid jargon and reflect the perspective, technical capacity and concerns of the public. The key points need to be constantly and consistently stressed. Another approach is to provide background information on any health effect involved in the cluster – for example, spontaneous abortions and birth defects – as well as information on the scientific method, before presenting the report of the study results (Curbow et al, 1994).

The investigator should be straightforward and honest about fact, speculation and what is known, remain co-operative and responsive, and be prepared to provide any additional information rapidly. The most important factor is for the investigator to be respected as a credible source of information.

The principles and techniques of risk communication and risk perception are comparatively new to epidemiology. Sandman (1991a) has offered epidemiologists eight guidelines for public communication.

- Tell the people who are most affected what you have found, and tell them first.
- Make sure people understand what you are telling them and what you think the implications are.
- Develop a mechanism to bolster the credibility of your study and your findings.
- Acknowledge uncertainty promptly and thoroughly.
- Apply epidemiological expertise where it is called for and do not misapply it where it is unlikely to help.
- Show respect for public concerns, even when they are not scientific.
- Involve the affected people in the design, implementation, and interpretation of the study.
- Decide that communication is part of your job, and learn its basic principles.
Appendix 1 - Cluster Report Form

Suspected Cluster Report

Name of person completing form .................................................................

Date .........../......./........

Informant

First name ................................ ................................ Surname

................................ ................................ .................................................................

Address ................................ ................................ .................................................................

Telephone no. .......................

Background of informant:

  media (Specify) ................................ .................................................................

  family member of index case(s) (Specify relationship) ..............................

  friend of index case(s) ................................ doctor

  other (Specify) ................................ .................................................................

Description of the problem:

................................ ................................ .................................................................

................................ ................................ .................................................................

How did the informant come to believe there might be a problem?

................................ ................................ .................................................................

................................ ................................ .................................................................
Setting of the suspected cluster:

- neighbourhood
- school
- workplace
- other (Specify)

Is there a suspected exposure?

- yes (Specify)
- no

Some of the following information about the index case(s) may be available from the informant:

**Index case** ....... (number eg, 1, 2 etc)

- First name
- Surname

- Current address (last if deceased)

- How long has the person lived there?

- Age
- Date of birth
- Sex: male female
- Ethnicity: NZ European NZ Maori Pacific Islander (Specify)
- Other (Specify)

- Diagnosis
- Basis of diagnosis
- Date of diagnosis
- Date of death
- Place of death
Suspected environmental exposures:

- type of exposure .................................................................
- address where exposure occurred if different from above address ..........
- date exposure began ..............................................................
- date exposure ended ..............................................................
- detailed of changes in exposure (eg, when, extent, duration) ............... 

Smoking history (year started, duration, amount/day, tobacco type (eg, cigarettes))

Occupational history:

<table>
<thead>
<tr>
<th>type of industry</th>
<th>job</th>
<th>year job began</th>
<th>year job ended</th>
</tr>
</thead>
<tbody>
<tr>
<td>present job</td>
<td></td>
<td>19</td>
<td></td>
</tr>
<tr>
<td>last job</td>
<td></td>
<td>19</td>
<td>19</td>
</tr>
<tr>
<td>job before that</td>
<td></td>
<td>19</td>
<td>19</td>
</tr>
<tr>
<td>job before that</td>
<td></td>
<td>19</td>
<td>19</td>
</tr>
</tbody>
</table>

Any other details from informant:

Is the informant willing to assist in providing further information if necessary?  
yes  no
Appendix 2 - Data Recording

The aim of using a data collection form is standardised information collection.

Data collection forms should be as useful as possible to any investigation and minimise error in measuring exposure. They make explicit whether certain information is missing (eg, from records), is unavailable, or has not been sought.

Types of data collection form are:

- suspected cluster report form/log
- data abstraction form for data collected from medical records
- questionnaire.

Data collection is the first basic step in evaluating a possible cluster. It is essential to gather enough information to decide on further action.

Data Abstraction Form

- Do not abstract data if its quality is suspect (eg, data is inconsistent within the record)
- Identify source (eg, hospital records, general practitioner records, cancer registry)
- Items likely to be grouped together in the record should be in the same section of the form
- To minimise overlooking information, record information on the form as it is found in the records rather than in the order on the form
- Abstract data from its primary source (eg, laboratory result slip) rather than secondary sources (eg, laboratory results in medical notes)
- If possible, enter information direct from records to a computer file (using a laptop) with logic and range checks which allow data to be corrected immediately. This removes one error prone step in data collection.
Questionnaire

- Include in the introduction: informed consent, name of the research organisation and interviewer, and statements about the research topic and purpose, future use of data and confidentiality
- Include a box for the subject’s identification number
- Include a box for the date the questionnaire is administered
- Make instructions clear
- Use different type faces for instructions, questions and responses
- Locate response codes against the right-hand margin to assist data entry
- Circling a code number simplifies data entry
- When measuring exposure(s), use open-ended questions as far as possible
- Provide boxes for coding open-ended questions
- For closed-ended questions, response categories should be simple, brief and cover the full range of relevant answers. If only one response category is to be selected, the categories should be mutually exclusive. If more than one response category could be selected, have “yes/no” responses for each category. If response categories are not exhaustive give a final open category (eg, “Other (Please give details)”)  
- Questions with few possible responses (eg, sex) should be precoded – ie, circle a number corresponding to the appropriate response category (“1 = female, 2 = male”). If there are many possible responses (eg, occupation), record in words with space for later coding.
- Make response codes consistent (eg, “1 = no, 2 = yes”, “8 or 88 = not applicable”, “9 or 99 = non-response/ not known”)
- Code sequences in the same numerical direction, consistently increasing from 1 – for example, “nil = 1, mild = 2, moderate = 3, severe = 4”, or “low = 1, medium = 2, high = 3”
- Provide for coding without losing information; ie, do not categorise continuous data at the coding stage
- Do not require the data collector to make calculations
- Design forms for direct data entry into a computer
• Obtain copies of questionnaires used previously by others

• Use standard questions where possible (eg, demographic questions from the Census)

• Decide what range of answers are most important to the investigation and then determine the questions which are likely to yield this

• Group questions in topics and proceed from the general to the particular within a topic

• Certain demographic questions (eg, ethnicity, income) are sometimes threatening and are more appropriately asked at the end

• Use a logical sequence – for example, proceed from the present to successively earlier time periods

• Cover all items needed to fully characterise the exposure(s) of interest – for example, nature, amount (dose), time relationships (start, duration, variation over time)

• A matrix format is useful for recording exposures (eg, occupations, place of residence)

<table>
<thead>
<tr>
<th>type of industry</th>
<th>job</th>
<th>year job began</th>
<th>year job ended</th>
</tr>
</thead>
<tbody>
<tr>
<td>present job</td>
<td>..........</td>
<td>................</td>
<td>19 .............</td>
</tr>
<tr>
<td>last job</td>
<td>..........</td>
<td>................</td>
<td>19 .............</td>
</tr>
<tr>
<td>job before that</td>
<td>..........</td>
<td>................</td>
<td>19 .............</td>
</tr>
<tr>
<td>job before that</td>
<td>..........</td>
<td>................</td>
<td>19 .............</td>
</tr>
</tbody>
</table>

• Choose the words carefully – avoid abbreviations and technical jargon

• Avoid vague descriptors such as “usually”, “regularly” – be precise when quantifying

• Do not include more than one concept in a question (eg, do not say “Have you ever used ... or ...?”)

• Use a specific time reference (eg, “Before [reference date], how often ...?”)

• Avoid ambiguous questions (eg, “Sex?” should be worded “What sex are you?”)

• Avoid leading questions
• Ask for actual details rather than asking the informant to make judgements that may increase recall error (e.g., record details of major changes in exposure rather than average annual exposures)

• Review each question and omit if it does not contribute to achieving the investigation’s objectives

• Have the draft questionnaire peer reviewed

• Pretest the questionnaire.
Glossary

Carcinogen
A substance capable of causing cancer.

Confidence interval
A range of values for a variable that has a specified probability of including the true value of the variable.

Confounding
A situation in which a measure of the effect of an exposure on risk is distorted because of the association of the exposure with other factor(s) that influence the outcome.

Designated officer
A medical officer of health, health protection officer, or other officer designated by the Director-General of Health under the Health Act 1956 or other legislation.

Endemic
The constant presence of a disease or infectious agent in a given population group or geographic area.

Epidemiology
The study of the distribution and determinants of health-related states or events in specified populations.

Hazard
A situation or event of potential harm to health.

Latency
The time period between exposure to a disease-causing agent and the appearance of the manifestations of the disease.

Person-years
The number of years that a person in a study population has been observed.

Poisson distribution
A distribution function used to describe the occurrence of rare events or to describe the sampling distribution of isolated counts in a time or space continuum.

Relative risk
The ratio of the risk of disease or death among the exposed to the risk of disease or death among the unexposed.

Risk
The probability of harmful consequences arising from a hazard.

Risk assessment
The characterisation of potential adverse effects of exposures to hazards.
**Risk communication**
An interactive process of exchange of information and opinions among individuals, groups, and institutions.

**Sensitivity**
A measure of the probability of correctly diagnosing a case, or the probability that any given case will be identified by a test.

**Specificity**
A measure of the probability of correctly identifying a non-diseased person.

**Standardisation**
A technique used to minimise the effects of differences in age when comparing populations.

**Surveillance**
Data collection to detect events or identify trends to initiate public health action.

**Teratogen**
An agent which produces birth defects in an embryo or fetus.
References


Bibliography

Cluster Theory


**Investigating Clusters**


Clusters in New Zealand


Investigating Clusters of Non-Communicable Disease


Communicating Risk


