# Epidemiologic Investigation

**UNIT 2**

## Introduction

*In most countries a routine health information system already exists, but health workers may have to take the time to uncover and analyse the “hidden” data it contains. This system will provide useful information or at least a good “guestimate”. Make the best use of the data being collected, rather than rejecting it as unsuitable. On the other hand, the collection of useless and unnecessarily complicated data should definitely be discouraged.* (Vaughan & Morrow, 1989, 34)

Unit 2 covers various aspects of epidemiologic investigation. Study Session 1 explores the various sources of epidemiological information. Issues surrounding the quality of the information and data sourced are examined. In Study Session 2, you will need to think about where to source information from your own area and learn to analyse and interpret this data. This will enable you to better understand the health status in your area and how to use health information for planning purposes.

In Study Session 3 you will work with historical and more recent data related to outbreaks of infectious diseases. This entails plotting information in graph and table form and interpreting the patterns that emerge.

The final two Study Sessions deal with screening and surveillance. Examples from a variety of screening and surveillance programmes are presented and you are required to explore the contextual, social and ethical issues that emerge. Throughout this unit, you will draw from your own experience and context. You will become aware of how crucial it is to have reliable health information and to conduct well-planned epidemiological investigations.

There are five Study Sessions in Unit 2.

Study Session 1: The Source and Quality of Epidemiological Information Study Session 2: Epidemiological Information

Study Session 3: Infectious Diseases and Outbreaks Study Session 4: Screening

Study Session 5: Surveillance

# Unit 2 - Study Session 1 The Source and Quality of

**Epidemiological Information**

## Introduction

The previous Study Sessions introduced different types of data which contribute to the process of health measurement in some way. These types of data include demographic data, information on health status, health services, disease frequency measures, and so on. You also discovered that the data you need is, unfortunately, not always easy to access. Inevitably, it is necessary to go through a variety of different sources of information to locate the data you need. For this reason, it is important that you know where to search for the best sources of information.

Once you have established the existence of the information you need, you then have to determine how to get access to the information and how to assess how reliable it is, i.e. its value and quality. These questions are the subject of this Study Session.

## Contents

1. Learning outcomes of this session
2. Readings
3. Identify useful sources of epidemiological information
4. Assess the value, quality and accessibility of health information
5. Session summary

## Timing of this session

This Study Session contains three readings and five tasks. It should take you up to three hours to complete.

## LEARNING OUTCOMES OF THIS SESSION

|  |  |
| --- | --- |
| **By the end of this Study Session you should be able to:** | |
| ***Health Measurement Outcomes***   * Identify key sources of epidemiological information. * Assess the quality, relevance and accessibility of community health data. | ***Academic Learning Outcomes***   * Collect information and data from a variety of sources. * Critically compare different types of information. |

1. **READINGS**

There are three readings to which you will be referred in the course of the Study Session.

|  |  |
| --- | --- |
| **Author/s** | **Publication Details** |
| Vaughan, JP & Morrow, RH | (1989). Ch 4 – Epidemiological Health Information. In *Manual of Epidemiology for District Health Management.* Geneva: WHO: 33–44. |
| Joubert, G., Ehrlich, R., Katzenellenbogen,  J. M. & Abdool Karim,  S. S. | (2007). Ch 13 – Routine health information systems and disease surveillance. In *Epidemiology: A Manual for South Africa.* 2nd Ed. Cape Town: Oxford University Press: 172–187. |
| Vaughan, J. P. & Morrow, R. H. | (1989). Ch 7 – Epidemiological Surveys. In *Manual of Epidemiology for District Health Management.* Geneva: WHO: 81–84. |

## IDENTIFY USEFUL SOURCES OF EPIDEMIOLOGICAL INFORMATION

The sources of information you decide to use will vary depending on the kind of work you are doing, your location in the health system and the specific questions you want the data to help you answer.

The following task enables you to examine a number of commonly used sources of information.

**READING**

Vaughan, J. P. & Morrow, R. H. (1989). Ch 4 – Epidemiological Health Information. In

*Manual of Epidemiology for District Health Management.* Geneva: WHO: 33–44.

**TASK 1 – Identify useful sources of information**

1. List 6 – 10 sources of information you might consider useful to your work. Give reasons for your choice.
2. Compare your list with the list discussed on page 35 of Vaughan & Morrow (1989) and the list included after the next task below.
3. Compare your reasons for the sources you chose with the comments presented by Vaughan and Morrow. What do you notice? Draw your own conclusions.

#### FEEDBACK

Some additional sources of information are discussed in Vaughan and Morrow on page

53. Under certain circumstances, it may become necessary to supplement the routine data collection system with one or other of these more specialised methods of obtaining information from these sources, e.g. investigations of outbreaks or surveys. Take a look at Chapter 13 of Joubert *et al.* (2007) which also provides an overview of routinely available data, and includes a few interesting examples of data collection practices and problems.

**READING**

Joubert, G., Ehrlich, R., Katzenellenbogen, J. M. & Abdool Karim, S. S. (2007). Ch 13 – Routine health information systems and disease surveillance. In *Epidemiology: A Manual for South Africa.* 2nd Ed. Cape Town: Oxford University Press: 172–187.

**TASK 2 - Assess the advantages and disadvantages of different sources of information**

1. Evaluate the advantages and disadvantages of 4 – 5 of the following sources of information: hospital clinic records, statistical data, data from other departments or organisations, health information systems, annual reports, internet, text books, research projects, district health profiles, journals, surveys, census data.

#### FEEDBACK

In a recent training programme with district health managers on health informatics, the group drafted a summary of their responses to this task in the table included below.

See how similar your own comments are to those recorded by this group.

**Advantages and disadvantages of different sources of data**

|  |  |  |
| --- | --- | --- |
| **ADVANTAGES** | **DISADVANTAGES** | |
| **HOSPITAL CLINIC RECORDS** | | |
| * Regular, monitored, fits into overall system. * Basic information re: community. * Information on individuals, families and communities. * Shows health trends. * There is a statutory obligation to keep records. |  | * Too much data, taking too long to collect. * Poor feedback and communication. * Very seldom analysed. * Single geographical area. * Information from health sector only. * Only as good as record-keeper. * Only clinic visits are covered. |
| **STATISTICAL DATA** | | |
| * Enables tracking of major trends. * Background information on broad social conditions. * Can go back to see trends. |  | * Not always accessible. * Causes of death not known. * Incomplete statistics. * Is not fed back to role-players. * Not always reliable. |
| **OTHER DEPARTMENTS OR ORGANISATIONS** | | |
| * Gives more comprehensive picture. * Gives inside story on community profile re: political power, cultural influences. * Can provide information on perceived community needs. |  | * Limited data / Not easy to access. * Unknown reliability. * Subjectivity (influenced by personal agendas). * Systems not compatible (age, standards). * Poor training in interpretation and use. |
| **HEALTH INFORMATION SYSTEM** | | |
| * Good source. * Linked up to other processes. * Can be interpreted locally. * Can empower people. |  | * Limited e.g. TB. * Questionable accuracy. * Needs intensive training of health worker. * Feedback and interpretation problematic. * Lack of incentive to analyse data locally. |
| **ANNUAL REPORTS** | | |
| * Profile of needs and provision of services. * More textured than statistics. * Gives direction to other sources. |  | * Poor analysis. * Biased and not always inclusive. * Too broad - summary. * Out of date due to lengthy production time. * Boring/not user friendly so not read. |
| **ADVANTAGES** | | **DISADVANTAGES** |
| **INTERNET** | | |
| * Huge source, instant access to e.g. literature. * Point of comparison e.g. international. |  | * Limited facilities for access. * Developing world less well represented. * Not always up to date! |
| **TEXT BOOKS** | | |
| * Expert information. * Identify broad range of problems. * Provides models/formats. |  | * Out of date. * Not always applicable. * Too much text. * Theory does not inform implementation. * Costs/availability. |

|  |  |
| --- | --- |
| **RESEARCH PROJECTS** | |
| * Current information. * Time saving (if relevant). * Can be specific to a problem. * Pilots/demonstration projects. | * Costly and unsustainable. * Not always applicable or relevant. * Takes time so not often up to date. * Author biased. * Creates expectations. |
| **DISTRICT HEALTH PROFILES** | |
| * Readily available. * Gives “bird’s eye” view. * Provides leads to other sources. | * Questionable accuracy and reliability. * Questionable validity of outdated profiles. * Takes time to keep updated. |
| **JOURNALS** | |
| * Current information. * Can get back numbers (old copies). * Summarised. * Different points of view. | * Difficult to access. * Can be biased. * More academic than practical. * High volume. |
| **SURVEYS** | |
| * Cost effective. * Focused/specific, pick up hidden information. * Can be spread over large areas e.g. national. * Correctly planned, much more comprehensive. | * Once-off information. * Influenced by questionnaire. * Might not empower people. * Create expectations. |
| **CENSUS** | |
| * Gives provincial/national picture. * Standardised denominators & baseline data. * Can compare with e.g. international data. * Useful for planning. | * Not available until years later. * Expensive. * Creates expectations. * Unknown reliability. |

In general, it is important to realise that inaccuracy and unreliability can lead to bad planning.

## ASSESS THE VALUE, QUALITY & ACCESSIBILITY OF HEALTH INFORMATION

Everyone will have slightly different data needs. What is important is to have a clear idea of why each item of data is important enough to collect and what you plan to do with it. This is the focus of the following task. Assessing the value of the data and checking for possible omissions should become a regular feature of your work.

**READING**

Vaughan, J. P. & Morrow, R. H. (1989). Ch 4 – Epidemiological Health Information. In

*Manual of Epidemiology for District Health Management.* Geneva: WHO: 33–44.

**TASK 3 – Critically assess a selection of data**

Critically assess a district health information data selection given by Vaughan and Morrow (1989) on page 43.

1. Decide what you would use each group of data for.
2. If you think certain data should be excluded from the list give your reason for this.
3. If you think certain essential groups of data have been omitted, identify this information and give the reason for adding it to this list.

#### FEEDBACK

If the information can be made available without too much difficulty and you have established a clear use for it, then its existence is probably justified. However it is also important to determine its quality, to establish that it is valid, up-to-date and reliable.

One of the hardest things to evaluate is the quality of the data that you receive.

At a later stage in the course, you will assess research data in terms of its validity and reliability. At this stage, it is important simply to determine whether the process of data collection and record-keeping is reasonably effective, for example, to check that the data is complete and up-to-date. This is the focus of the next task.

**TASK 4 – Evaluate the quality of sample data**

1. Check the information in the following table. Use a calculator where necessary.
2. Think of reasons that could account for any missing data.
3. Comment on the quality of the data collection process.

## Sample data on four patients

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Name** | **Patient 1** | **Patient 2** | **Patient 3** | **Patient 4** |
| **Record No.** | 6781 | 346 | 2487 | 2351 |
| **Age** | 785 | 69 | 52 | 08 |
| **Gender** | F | F | F | M |
| **Weight** | 85 | 70 | 80 | 105 |
| **Height** | 1.6 | – | 1.7 | 2.0 |
| **BMI** | 33 | 24 | 38 | 26 |
| **Blood pressure 1** | 150/110 | 135/95 | 130/95 | 160/110 |
| **Blood pressure 2** | 140/100 | 125/85 | 120/85 | 150/100 |
| **PAP smear result** | Neg. | Pos. | Neg. | Pos. |
| **Breast lump check** | Neg. | – | Neg. | Neg. |
| **Urine sugar** | Pos. | – | – | Pos. |

#### FEEDBACK

If a patient’s weight is recorded as 650 kg, the chances are that someone made a mistake when writing down 65 kg. If the number of patients attending a clinic suddenly rises from an average of 500 per week to over 2 000, you need to find out what happened. The chances are that someone has added another set of data that you do not yet know about, or they simply sucked the figures out of their thumb! Or you need to ask whether the rules governing data collection have undergone a sudden change.

In the table above, there are several gaps. This may mean that data was not entered, or that the procedures from which the data arises were not carried out. It might also be that the data is irrelevant, for example a breast lump check in a male patient. The information might be written in the wrong place, such as a PAP smear result in the records of a male.

Digit preference can occur, for example rounding off to the nearest 0 or 5 in all the weight and BP readings. There may have been eagerness to show improvements in follow-up readings. Note that all the BP2 readings are exactly 10 lower than those in BP1.

Typing an extra digit makes a nonsensical age entry of 785 for Patient 1. A digit switch gives an age of 08 instead of a more likely 80 for Patient 4 who is clearly adult age and size. A calculation or typing error enters a BMI of 38 instead of 28 for Patient 3.

The previous tasks suggest that a data collection process in any health facility needs to be goal-directed, evaluated, clearly structured and known by all participants. The implications of any changes of procedure should be recognised and clearly communicated to anybody who uses the data.

**TASK 5 - Assess the data collection process where you work**

Select one type of health data collected in the health facility where you work. Investigate the process by which it is collected, recorded and summarised. Look at your data sample and comment on its quality by asking these questions, and any others that you feel are relevant:

1. Is the data all there?
2. Is it reliable?
3. Is it true data or has it been fabricated?
4. Despite its limitations, can it still be useful?

#### FEEDBACK

You might like to review the definitions of variables, repeatability and validity on pages 81–84 of the following reading. While these particular comments are more directly related to the collection of research data for surveys, they also have some relevance to the routine data collection process. Do not concern yourself too much with the more technical calculation of reliability at this stage.

**READING**

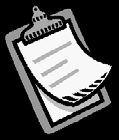
Vaughan, J. P. & Morrow, R. H. (1989). Ch 7 – Epidemiological Surveys. In *Manual of Epidemiology for District Health Management.* Geneva: WHO: 81–84.

A major objective of your data collection process is to ensure that data is adequate for decision-making. Is it good enough to help you assess what is going on? If the data collection process appears to be functioning properly and the dataset is complete, the information is probably good enough to use. Even if the quality is deficient in certain respects, you might still be able to use it if you know in which way this deficiency will actually affect its interpretation. For example, will a small error in the recording of repeat visits to your clinic lead to an underestimate or an over-estimate of the total patient attendance figures? (It gives you a small overestimation of patients served by the clinic).

## SESSION SUMMARY

In this Study Session, we examined different sources of information and considered how poor or erratic data collection procedures can distort findings. Your attention was drawn to questions that you should ask about any dataset.

The next Study Session explores several different sources of epidemiological data and the importance of goal-directed data collection is discussed. You will gather data from your own context and learn to conduct a mini-situation analysis.



Every source of data has its advantages and inconveniences and combining multiple sources of data can be useful.

The source of data will be selected according to the purpose of the data collection or research: the source will differ according to whether one needs exhaustive information or to obtain data very quickly.

When establishing a data collection tool, the most important step is to decide WHAT we want to collect and WHY.

The quality of data has to be assessed systematically using logic, e.g. Are there outliers? Are there missing values? Are there strange values when taken as a set (repetition for instance)? Are there absurd values?

**KEY POINTS IN UNIT 2 SESSION 1**

**The Source and Quality of Epidemiological Information**

# Unit 2 - Study Session 2 Epidemiological Information

## Introduction

Most of us have encountered health and heath measurement information in various forms. However, most Public Health practitioners have experienced great difficulty in accessing relevant information they can use in their work. If the information does exist, they may discover enormous amounts of information of very poor quality.

As a Public Health worker, it is inevitable that you will be involved either in recording and organising health information, or in summarising and interpreting existing sets of data in order to make decisions. Knowing the different types of data that are used, where they can be found, and assessing their quality are, therefore, important aspects in this process and form the focus of this Study Session.

In the course of this Study Session, you will need to spend some time thinking about where you would search for local data in order to feed into and influence your own work

. Knowing where to find useful information and being able to critically assess its value are essential skills for Public Health workers.

## Contents

1. Learning outcomes of this session
2. Readings
3. Sources of epidemiological data
4. Factors that influence data collection
5. Recognising data inaccuracies and gaps
6. Compare local and global health data
7. Session summary

## Timing of this session

There are three readings, three tasks in this Study Session.

## LEARNING OUTCOMES OF THIS SESSION

|  |  |
| --- | --- |
| **By the end of this Study Session you should be able to:** | |
| ***Health Measurement Outcomes***   * Assess the quality, relevance and accessibility of community health data. * Review the contents of a basic district health dataset. * Interpret commonly reported epidemiological indicators. | ***Academic Learning Outcomes***   * Locate different sources of (health) information. * Compare different indicators in published (health) datasets. * Interpret data in terms of its context. |

1. **READINGS**

|  |  |
| --- | --- |
| **Author/s** | **Publication Details** |
| Health Systems Trust | (1996). *How to Conduct a Situation Analysis: A Guide for Health Districts in South Africa.* Durban: HST: 1–8 & 50–53. |
| Vaughan J. P. & Morrow, R. H. | (1989). Ch 1 – District Health Management. In *Manual of Epidemiology for District Health Management.* Geneva: WHO: 1–8. |
| UNICEF | (2014). Basic Indicators. In *The State of the World’s Children 2014 In Numbers: Every Child Counts.* Geneva: UNICEF: 42-47. |

1. **SOURCES OF EPIDEMIOLOGICAL DATA**

For anyone working in the Public Health field, it is important to have easy access to the wealth of information available, although as we have said, it is often inaccurate and irrelevant.

Previous activities and readings, e.g. Chapter 1 in Bonita, Beaglehole & Kjellstrom (2006), have illustrated the many different types of information or data used in epidemiology and the wide variety of sources from which they may be obtained.

**READINGS**

Health Systems Trust. (1996). How to Conduct a Situation Analysis: A Guide for Health Districts in South Africa. Durban: HST: 50–53.

Vaughan, J. P. & Morrow, R. H. (1989). Ch 1 – District Health Management. In

*Manual of Epidemiology for District Health Management.* Geneva: WHO: 1–8.

Health System Trust’s 1996 manual lists some useful sources of information from page

50. On page 4, Vaughan & Morrow (1989) also summarise the kinds of information that may be useful. These lists may help you identify data that is relevant to your own work and to compile your own list of useful sources on an ongoing basis, adding new sources as you find them. Becoming familiar with sources of data for your own district or area is crucial to being able to work effectively in this field.

## FACTORS THAT INFLUENCE DATA COLLECTION

Effective data collecting should be as goal-directed as possible. That is, you should identify the specific questions that you wish to answer, and then search for the best possible sources of information to answer these questions. The sources you consult will vary with the type of information required, the context and purpose of data collection, e.g. to establish the cause of a health problem.

The role that health information plays at the local (community or district) level can be very different from the role it plays at provincial or national level. The way you collect, interpret and later apply the data to activities in your workplace can also be very different for rural and urban areas, and for affluent and poorer communities in your country. It is important to understand the context within which you will collect and work with the data, and how the context or environment influences this data.

#### The Influence of Context and Purpose

Study Chapter 1 of Vaughan and Morrow (1989) in which they approach the issue of data collection from the perspective of a district health officer who needs to know how to plan and manage the affairs of a district health system. It is important for you to be clear about *your* reasons for collecting health measurement information in your own work and context.

**READING**

Vaughan, J. P. & Morrow, R. H. (1989). Ch 1 – District Health Management. In *Manual of Epidemiology for District Health Management.* Geneva: WHO: 1–8.

**TASK 1 – Explore factors that influence data collection**

As you read the above text, try to answer the following questions.

1. What are the main purposes for collecting information in the district health context? See page 4 of Vaughan and Morrow (1989).
2. How does this compare with the data that you would need in your own context?

How does context affect the data collection process?

1. According to this text, what are the two main sources of health information?
2. Why are these two types of information considered so important?

#### FEEDBACK

1. It has been made clear that the data collected as well as the data collection process must be in line with the reasons for wanting it in the first place. Think how such data could be used by the district health management team in carrying out its responsibilities as listed on page 3 of Vaughan & Morrow.
2. From the discussion on the health information needs of the district, you will have noticed that the circumstances in each district or community are unique. Slightly different risk factors may be present, and the community may experience different levels of ill-health. These unique local circumstances require interventions customised to match local Public Health needs. The importance of having accurate, locally relevant data with which to work should be evident.
3. Health information can be collected using a surveillance system or reports and surveys.
4. They are important because they allow us to combine the continuity of routine data collection (a surveillance system), which is used to monitor trends in disease and health over time, with the detail that special surveys or investigations can provide on specific health problems.

Until the recent introduction of the District Health Management Information System (DHMIS) in South Africa, government health services collected vast amounts of data on health and health care activity. It was an entrenched administrative ritual in which many productive hours were lost. The private sector has excelled in collecting itemised health care billing information, but remains incapable of reflecting the health status of the community at large. Indicators that were inappropriate and unable to really tell us what was going on in the health system were collected uncritically but never used. On the other hand, such large volumes of data were collected that the task of analysing and interpreting it for local use was simply too difficult. The result was that very little data ever got used in management decisions or in the ongoing planning and evaluation of district health activities.

It is therefore essential to approach the process of data collection or review with a very clear, specific purpose. Ensure that the questions you want answered by this information (your epidemiological purpose for data collection) are clearly defined in your own mind before you collect any information.

It is also important to be critical of all data you receive as it may contain inaccuracies, gaps or may even be irrelevant. Sometimes though, the gaps can give you really important information or insights into the nature of the health problem you are investigating.

#### Understand the Health Status of Communities

Another factor that influences data collection is the need to understand the health status of *whole communities*, rather than just the health of individuals. Broader social or geographical information may assist in the process of planning and implementing treatment or Health Promotion strategies, or evaluating their outcomes.

The practical application of this concept is described as a “situation analysis” or “community diagnosis”, as it is called in Vaughan & Morrow (1989), page 5.

**READINGS**

Health Systems Trust. (1996). *How to Conduct a Situation Analysis: A Guide for Health Districts in South Africa*. Durban: HST: 50–53.

Vaughan, J. P. & Morrow, R. H. (1989). Ch 1 – District Health Management*.* In *Manual of Epidemiology for District Health Management.* Geneva: WHO: 5.

**TASK 2 - Epidemiology, community diagnosis and the planning cycle**

1. What do you think is the main purpose of carrying out a “community diagnosis”?
2. Why does it include information that is not specifically about health and health care?
3. How does this kind of diagnosis differ from the clinician’s diagnostic process?

4 Both the readings above refer to the *planning cycle*. What do you believe is the role of epidemiology in this process?

5. Where, in this cycle, do you think the community diagnosis might occur?

#### FEEDBACK

As illustrated in Vaughan & Morrow (1989) page 6, Table 1.1, the rationale for the community diagnosis is similar to that of a clinician trying to understand a patient’s problem and prescribing treatment. The main difference is that where the clinician is concerned with the individual subject, in community diagnosis the community is the collective subject of the diagnostic process. The non-health information is vital as it has the potential to describe the major environmental influences or determinants that have the capacity to increase or decrease the risk of illness in the community.

Refer to your definition of epidemiology from Study Session 1: you might now conclude that epidemiology is only concerned with health data, *or* that it embraces the whole process of combining different types and sources of information that help us understand a health problem. Revise your definition if your understanding has changed.

In carrying out a situational analysis, you need to select a relevant set of variables that will enable you to conduct a community diagnosis. In the next section, we look briefly at data inaccuracy issues.

## RECOGNISING INACCURACIES AND GAPS

Those of you who collect data regularly will be familiar with the kinds of gaps and inaccuracies which are likely to occur. You will also probably be aware of the implications of poor data in health systems management.

Here are some of them:

* If data is grossly inaccurate then managers, supervisors and staff will not use the information generated from the data.
* If people do not use the information, then all the effort and time that large numbers of people spent collecting the data and developing the information system will have been wasted.
* If the data is grossly inaccurate, but people do not realise this, and they use the information for planning and general decision-making, then it is likely that the plans will be flawed and poor decisions will be made, because they were based on inaccurate data.

Inaccurate data, if used in decision making, can therefore be very harmful, giving rise to unnecessary problems.

Errors can be detected using *General Accuracy Checking Measures*. These measures are things that are well known and just common sense to do.

## General Accuracy Checks

|  |  |
| --- | --- |
| **Check for completeness** | Are there any gaps, missing data? |
| **Check proper placing** | Is the data captured in right box, or is data entered wrongly due to typing errors, e.g. unlikely valuables for variable such as a male being pregnant, or low birth weights exceeding the number of deliveries. |
| **Check arithmetic** | Are there any mistakes in adding or calculating data such as Body Mass Index (BMI)? |
| **Identify contradictions between variables** | Is there any data that is unlikely or does not make sense, e.g. 100 births in a month in a community where there are only 2 000 women of child-bearing age. |

A good quality dataset should be correct, complete and consistent. It should be reliable and accurate enough to support decisions. Finally, it should be comparable i.e. using the same definitions. Ask whether the dataset you have been provided with complies with all these criteria.

The next section will provide an opportunity to compare some of these indicators between countries.

## COMPARE LOCAL AND GLOBAL HEALTH DATA

In your next task you will examine similar health indicators drawn from a number of national and international datasets.

This has two purposes. The first is to see the kinds of indicators considered appropriate for publication in such global datasets. The second is to enable comparison of health status across different continents and countries globally. You will be able to rank your own district or country in relation to others for which data is listed.

**READING**

UNICEF. (2014). Basic Indicators. In The State of the World’s Children 2014 In Numbers: Every Child Counts. Geneva: UNICEF: 42-47.

UNICEF. (2014). Statistical Tables. In The State of the World’s Children 2014 In Numbers: Every Child Counts. Geneva: UNICEF: 21-35.

**TASK 3 - Work with tables of international health indicators**

In this task, you need to use the tables from the 1998 UNICEF *State of the World’s Children* Report which are provided in the Reader. You will make comparisons of data between rich and poor countries. Choose 2 or 3 countries.

Examine and answer these questions in relation to these countries:

1. Begin by scanning the title of each column to identify what data is included and what type of measurement is used to quantify it.
2. Scan down the columns to get an idea of the range of the reported values, i.e. the largest and smallest figures. Also note whether there are any missing (or unreported) indicators. Why do you think this occurs? Does it influence any comparisons you may wish to make?
3. Look up any terms that are new to you in the glossary contained in Vaughan and Morrow (1989), Ch 14, page 155.
4. In the UNICEF Basic Indicators tables, select a country that interests you and make a note of the Under-5 mortality rate. How has this changed from 1960 – 1996?

1. How does this country’s Under-5 mortality rate compare with that of a much richer country and of a much poorer country?
2. How does this country compare with your own country or district?
3. Do any of the other indicators listed help to explain these differences between countries?
4. Does the Infant Mortality rate differ in the same way as the Under-5 mortality rate?
5. Which economic variable can tell you about the poverty or wealth status of each country?
6. Does there appear to be an association between poverty and mortality rates?
7. What is the relevance of including Adult Literacy, % Household Income and Primary School Enrolment in such a table?
8. Examine the Health table in the same way. Study *Measles Immunisation* and

*Access to Safe Water* as the basis for comparison of the same 2 – 3 countries.

#### FEEDBACK

It is important to recognise that even the impressively large and well-organised WHO and UNICEF datasets are built up using summarised national data. This in turn was probably compiled from the same problematic district or regional data discussed in the previous task. A further complication of international datasets is the need to ensure that all reports use the same standardised definitions of the indicators. For example, we found that the calculation of maternal mortality was done differently in two published datasets. This makes comparison difficult. It is sometimes unclear whether the data has been collected over exactly the same periods. You also need to remember that all the figures reported are national averages, and they therefore hide any variations that may exist between groups within countries.

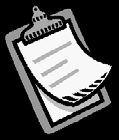
It is likely that as you compare richer countries, i.e. countries with high Gross National Product (GNP), with poorer ones, i.e. countries with low Gross National Product (GNP), you will notice that certain health indicators appear to be better in the more wealthy countries. However GNP is a very crude measure of the economic status of a country. While it tells you about the overall national income, it does not reflect the skewed distribution of income between a few extremely affluent people and the many more poor people in the population.

Consequently other more specific financial or social indicators may be more meaningful, such as household income or female literacy. Although these may not be health indicators in the narrow sense of the term, they are helpful in association with certain health status indicators. Why do you think this is so?

1. *SESSION SUMMARY*

This Study Session explored several different sources of epidemiological data and emphasised the importance of goal-directed data collection. . Competence in, and an understanding of the collection of data from your own context and evaluating different indicators and comparing data from national and international datasets is essential for work you will do in this field.

In the next Study Session we move away from routine data collection to discuss outbreaks and infectious diseases.



Data collection should be goal-orientated, meaning that before starting data collection, you need to have in mind the set of questions to which you want an answer.

By understanding the context (socio-economic, historical, geographical, political, etc) of your data (and then of your population), you are already making a step towards identifying the determinants of health of this population. You will also have better insight into planning health interventions if you already have information about the broader context of the population.

While analysing, you need to be critical towards your data and should not “believe” them at the first sight. A good quality dataset should be correct, complete, consistent, accurate and comparable.

**KEY POINTS IN UNIT 2 SESSION 2**

**Epidemiological Information**

# Unit 2 - Study Session 3 Infectious Diseases and Outbreaks

## Introduction

In many respects, it was with the investigation of infectious diseases that the discipline of epidemiology actually began. Just over a century ago, the so-called father of modern epidemiology, John Snow, mapped a number of cholera outbreaks that occurred in England. You have read through some of the data from this outbreak in Chapter 1 of Joubert *et al.* (2007). Although the plague, typhus, malaria and numerous other conditions had been around for some time, they had not been systematically recorded and analysed before. Another early practitioner of epidemiology was Florence Nightingale, a nurse, who after a systematic analysis of death and disease patterns, made dramatic changes to the risk of death from disease by soldiers in the Crimean war.

In this Study Session, a number of examples of infectious disease outbreaks are examined. They serve to illustrate some of the main characteristics of epidemics, some of the terminology associated with epidemiology, and various types of graphical representations used to display epidemiological data.

## Contents

1. Learning outcomes of this session
2. Readings
3. Define and clarify new terms
4. Describe an outbreak or epidemiological event
5. Session summary

## Timing of the session

This Study Session contains two readings and five tasks. It should take you about two and a half hours to complete.

## LEARNING OUTCOMES OF THIS SESSION

|  |  |
| --- | --- |
| **By the end of this Study Session you should be able to:** | |
| ***Health Measurement Outcomes***   * Define and measure infectious disease events in communities. * Analyse data from an outbreak of infectious disease. | ***Academic Learning Outcomes***  Interpret a set of data.  Illustrate a set of data using simple graphical representation techniques. |

* 1. **READINGS**

There are two readings for this Study Session.

|  |  |
| --- | --- |
| **Author/s** | **Publication Details** |
| Vaughan, J. P. & Morrow, R. H. | (1989). Ch 6 – Controlling an Epidemic. In *Manual of Epidemiology for District Health Management.* Geneva: WHO: 59–70. |
| Bonita, R., Beaglehole, R. & Kjellstrom, T. | (2006). Ch 7 - Communicable Diseases: Epidemiologic surveillance and response. Epidemiology. In *Basic Epidemiology.* 2nd Ed. Geneva: WHO: 117–122, 126-130. |

## DEFINE AND CLARIFY NEW TERMS

As always, familiarising yourself with key terms enables you to work more effectively with new information. Use Chapter 6 of Vaughan & Morrow (1989) to clarify the terms for Task 1.

**READING**

Vaughan, J. P. & Morrow, R. H. (1989). Ch 6 – Controlling an Epidemic. In *Manual of Epidemiology for District Health Management.* Geneva: WHO: 59–70.

|  |  |  |
| --- | --- | --- |
| **TASK 1 – Define new terms**  Define the following terms using your own words. Where possible, include an example to illustrate the meaning. | | |
| epidemic | reported cases | diagnostic criteria |
| active case detection | source of outbreak | epidemic curve |
| point source epidemic | propagated epidemic | spot map |
| attack rates | pandemic | endemic |
| **FEEDBACK** |  |  |

All of these terms are mentioned in Chapter 6 of Vaughan & Morrow (1989) or are listed in the glossary in Chapter 14 of the same manual. Hopefully, you have used your own words to explain the terms and tried to identify an example to illustrate the meaning of each term. Keep your explanations where you can easily refer to them.

The next task offers you a chance to develop a deeper understanding of the terms

*epidemic* and *endemic*. Use Chapter 6 of Vaughan & Morrow (1989) to assist you.

**TASK 2 - Determine whether a high level of disease is an epidemic**

Read the information and answer the questions.

1. In a number of poorer areas of Cape Town, the TB prevalence is close to 700 per 100 000, and it has remained this high for the past 10 years. But we are told it was not a common disease in Cape Town before that time. In Johannesburg, the TB prevalence does not exceed 200 per 100 000 in any suburb. In Durban, it has not exceeded 50 per 100 000 for years.
   * Is there a TB epidemic in Cape Town?
   * Is there an epidemic in Johannesburg?
2. In all of Mozambique and in a number of small towns in rural Mpumalanga, the annual incidence of malaria has been close to 40 per 100 000 for the past 20 to 30 years. Elsewhere in South Africa the incidence of malaria rarely gets closer than about 1 per 100 000.
   * Is malaria an epidemic or is it endemic to Mozambique and parts of Mpumalanga?

#### FEEDBACK

1. Review your definitions of the terms *epidemic* and *endemic*.

While the final conclusion remains somewhat debatable, there is a strong argument to suggest that Cape Town is experiencing a prolonged *epidemic* of TB since it was not always present at such high levels.

1. Mpumalanga and Mozambique, on the other hand, appear to have always had a substantial number of cases of malaria. The conclusion here is that malaria is *endemic* to this region.

Having clarified when a disease event is considered an epidemic, the next section introduces some of the processes we use to describe an outbreak of disease.

## DESCRIBE AN OUTBREAK OR EPIDEMIOLOGICAL EVENT

The purpose of the next task is to interpret and graphically represent two sets of disease outbreak data. The first set is from the early history of epidemiology, and uses notes made by John Snow during the 1850 cholera outbreak in London, a time when the causative agent for cholera was still unknown.

The second set of data is from more recent times. It is the Kaposi sarcoma data from New York during the early years of what later became the AIDS pandemic.

**TASK 3 – Plot and interpret epidemiological data**

1. First look at the data from Snow’s table of information on the London cholera outbreak in 1850. Use the data to give you an initial idea of what the outbreak was like.
2. Plot the data given for the two variables in Snow’s table on the set of gridlines supplied. Use a coloured pencil to plot the deaths and another colour for the onset of illness. Join the dots to form two line graphs.
3. What do the shapes of these *epidemic curves* suggest about the nature of the epidemic?
4. What does the data tell you about the duration of the illness?
5. If you assume that the handle of the affected water pump was removed on the 7th of September, what impact did that have on the epidemic?

**John Snow’s Table of Data on the London Cholera Epidemic**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Frequency of illness and deaths by date in the 1850 London cholera outbreak** | | | | | | | | |
| **Month** | **Date** | **Number of attacks** | **Number of deaths** |  | **Month** | **Date** | **Number of attacks** | **Number of deaths** |
| August | 27 | 1 | 1 | September | 11 | 5 | 15 |
| August | 28 | 1 | 0 | September | 12 | 1 | 6 |
| August | 29 | 1 | 1 | September | 13 | 3 | 13 |
| August | 30 | 8 | 2 | September | 14 | 0 | 6 |
| August | 31 | 56 | 3 | September | 15 | 1 | 8 |
| September | 1 | 143 | 70 | September | 16 | 4 | 6 |
| September | 2 | 116 | 127 | September | 17 | 2 | 5 |
| September | 3 | 54 | 76 | September | 18 | 3 | 2 |
| September | 4 | 46 | 71 | September | 19 | 0 | 3 |
| September | 5 | 36 | 45 | September | 20 | 0 | 0 |
| September | 6 | 20 | 37 | September | 21 | 2 | 0 |
| September | 7 | 28 | 32 | September | 22 | 1 | 2 |
| September | 8 | 12 | 30 | September | 23 | 1 | 3 |
| September | 9 | 11 | 24 | September | 24 | 1 | 0 |
| September | 10 | 5 | 18 | September | 25 | 1 | 0 |

**Line graphs showing cholera attacks and deaths 27 August to 25 September 1850**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Number of cholera attacks/deaths** | 150 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 140 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 130 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 120 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 110 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 100 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 90 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 80 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 70 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 60 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 50 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 40 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 30 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 20 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 10 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 0 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| **Date** | | 27 | 28 | 29 | 30 | 31 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 |
| **August** | | | | | **September** | | | | | | | | | | | | | | | | | | | | | | | | |
|  | | **Date of attack/Date of death** | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

|  |  |
| --- | --- |
| **Key** | Sample of colour used for each line graph |
| Attacks of cholera |  |
| Deaths due to cholera |  |

**FEEDBACK**

1/2. Compare your graph to the one on page 121 of Chapter 7 in Bonita, Beaglehole & Kjellstrom (2006).

As you first look at the data, you will see that the numbers of people who get sick or die each day are very low initially. Over a short period from 30 August until the 8 September, the numbers of people getting sick or dying suddenly increase substantially but subsequently quickly return to the same low levels evident prior to 30 August. You will also notice that almost exactly the same pattern is evident for both the illness-onset (attack) data and the death data.

**READING**

Bonita, R., Beaglehole, R. & Kjellstrom, T. (2006). Ch 7 – Communicable Diseases: Epidemiology surveillance and response. Epidemiology. In *Basic Epidemiology.* 2nd Ed. Geneva: WHO: 117–122, 126-130.

1. If you joined each dot plotted on the graph, you will have constructed two line graphs. They are almost identical in shape but separated by a period of about one day. This is roughly how long it must have taken for the disease to progress from onset to death. As you also noticed in the pattern of the raw data, the graph shows a steep rise and a sudden drop during the period from 30 August to

8 September. This is a classic epidemic curve, rising rapidly above the normally encountered prevalence of the disease and then returning to the original low levels thereafter.

1. You might ask why the number of cases suddenly decreased. Did the epidemic suddenly disappear? Perhaps the source of the epidemic was removed. Perhaps the rest of the neighbourhood fled to other parts of the city to avoid the disease. Perhaps almost everyone who could get the disease had already died and there were no people left who could get sick.
2. Removing the handle on the water pump on 7 September did not stop the epidemic. The number of cases had almost completely declined to pre-epidemic levels by the 7th. Removing the handle could have prevented a new group of people, who returned to the neighbourhood some time later, from drinking the water and thus preventing a new epidemic. Another steep epidemic curve could then have been plotted from the time this new group of people started to get ill.

Now plot another set of data. This time use a *histogram* (a bar chart) in which the height of each bar represents the number of cases of Kaposi sarcoma.

**TASK 4 – Plot more data**

1. The data obtained from Kaposi sarcoma notifications in New York for the period 1973–1982 is provided below. Use the data to draw a histogram in the diagram provided. The height of each bar should represent the number of cases of Kaposi sarcoma.

1. Compare the bars on the right to those on the left and comment on the general shape created. What does this suggest about the duration of the epidemic?
2. How does this compare with the data from the London cholera epidemic?

### Kaposi Sarcoma Notifications in New York: 1973 - 1982

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Year** | 1973 | 1974 | 1975 | 1976 | 1977 | 1978 | 1979 | 1980 | 1981 | 1982 |
| **Cases** | 0 | 0 | 0 | 0 | 2 | 2 | 4 | 4 | 30 | 85 |

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Number of notifications** | 90 |  |  |  |  |  |  |  |  |  |  |
| 80 |  |  |  |  |  |  |  |  |  |  |
| 70 |  |  |  |  |  |  |  |  |  |  |
| 60 |  |  |  |  |  |  |  |  |  |  |
| 50 |  |  |  |  |  |  |  |  |  |  |
| 40 |  |  |  |  |  |  |  |  |  |  |
| 30 |  |  |  |  |  |  |  |  |  |  |
| 20 |  |  |  |  |  |  |  |  |  |  |
| 10 |  |  |  |  |  |  |  |  |  |  |
| 0 |  |  |  |  |  |  |  |  |  |  |
|  | | 1973 | 1974 | 1975 | 1976 | 1977 | 1978 | 1979 | 1980 | 1981 | 1982 |
| **Year in which notification occurred** | | | | | | | | | |

#### FEEDBACK

1. As with the cholera data, start by looking at the raw data. Quite evidently the incidence of Kaposi sarcoma is very low initially and rises rapidly towards the end of the ten-year period represented here. Compare your graph to the one found on page 98 of the following reading.

**READING**

Bonita, R., Beaglehole, R. & Kjellstrom, T. (2006). Ch 7 – Communicable Diseases: Epidemiology surveillance and response. Epidemiology. In *Basic Epidemiology.* 2nd Ed. Geneva: WHO: 117–122, 126-130.

1. In this chart, the bars toward the right are much taller than those on the left. This resembles the beginning of the epidemic curve drawn with the cholera data. The shape suggests this is the beginning of an epidemic and it is unclear when it might eventually start to turn around and return to the low levels of 1973.
2. As you may be aware, Kaposi sarcoma is one of the clinical outcomes of AIDS, but this relationship was unknown in the early days of the AIDS epidemic. This is similar to the 1850 cholera epidemic, when vibrio cholerae, the causative agent, had not yet been identified.

The next task gives you the opportunity to explore different strategies for dealing with an epidemic should one occur in your area.

**TASK 5 - Select strategies for control of Snow’s cholera epidemic**

Assume the same epidemic recorded by Snow occurred today in your area. List briefly how you would:

1. target the source of the epidemic
2. interrupt transmission
3. protect susceptible people.

#### FEEDBACK

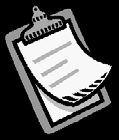
Compare the suggestions you listed with those in Table 6.2 on page 67 of Vaughan, J.

P. & Morrow, R. H. (1989). Ch 6 – Controlling an Epidemic. In *Manual of Epidemiology for District Health Management.* Geneva: WHO.

## SESSION SUMMARY

Some of the earliest experiences in epidemiology arose from efforts to map epidemics of disease in the mid-19th century. By mapping the same data, we have illustrated some of the most prominent features of disease outbreaks, and helped to define key terms such as *endemic* and *epidemic*. The Study Session also utilised several different types of graphical illustrations.

The next Study Session is concerned with the principles and role of screening as a part of disease control. You will examine the general advantages and disadvantages of different screening programmes and their effectiveness in preventing disease.



The purposes of an outbreak investigation are to find the source of the epidemic in order to stop the epidemic, and for further protecting the population susceptible to the risk.

**KEY POINTS IN UNIT 2 SESSION 3**

**Outbreak Surveillance**

**Self-evaluation question:** Based on what you learned so far, imagine situations where you are facing a false outbreak due to the quality of the data. How could this happen?

# Unit 2 - Study Session 4 Screening

## Introduction

Screening is an essential part of disease control. It involves the examination of apparently healthy people - i.e. people who do not have the symptoms of the disease (asymptomatic) - in order to classify them as likely or unlikely to have the disease.

Screening does not serve to diagnose; but rather identifies people who are likely to have the disease, for further investigations. Screening is also used as a prevention measure, to identify people who are *at risk* of developing a certain disease.

Screening is most commonly and effectively used to measure blood pressure, cholesterol and glucose levels. It uses a variety of tools from x-rays (e.g. mammography for breast cancer) and clinical procedures (e.g. colonoscopy for colon cancer) to blood and urine tests (for diabetes, HIV, and so on).

This Study Session introduces you to the key aspects and principles of screening and guides you how to assess the advantages, disadvantages and the accuracy of screening tests and programmes.

## Contents

1. Learning outcomes of this session
2. Readings
3. Define new terms
4. Principles of screening
5. Interpret screening results
6. Reflect critically on the practice of screening
7. Session summary

## Timing of this session

This Study Session contains two readings and five tasks. It should take you about two hours to complete.

## LEARNING OUTCOMES OF THIS SESSION

|  |  |
| --- | --- |
| **By the end of this Study Session you should be able to:** | |
| ***Health Measurement Outcomes***   * Describe the role of screening in health care. * Apply criteria for the introduction of screening programmes. * Interpret the results of screening tests. | ***Academic Learning Outcomes***   * Define and apply new terminology. * Consider intention and goals in order to select from a range of instruments. * Examine the advantages and disadvantages of a process. * Apply a set of pre-determined criteria. * Critically assess a screening programme. |

1. **READINGS**

There are two readings in this Study Session. You will be referred to them where relevant.

|  |  |
| --- | --- |
| **Author/s** | **Publication Details** |
| Bonita, R., Beaglehole, R. & Kjellstrom, T. | (2006). Ch 6 – Epidemiology and Prevention. In *Basic Epidemiology.* 2nd Ed. Geneva: WHO: 110-114. |
| Morrison, A. S. | (1998). Ch 25 – Screening. In *Modern Epidemiology (2nd edition).*  Philadelphia: Lippincott Williams & Wilkins: 499–510. |

## DEFINE NEW TERMS

As with each of the other areas of health measurement, screening also has a number of new terms and concepts to describe different approaches of testing in health care. The task that follows clarifies these terms.

**READINGS**

Bonita, R., Beaglehole, R. & Kjellstrom, T.

(2006). Ch 6 – Epidemiology and

Prevention. In Basic Epidemiology. 2nd Ed. Geneva: WHO: 110-114.

Morrison, A. S. (1998). Ch 25 – Screening. In Rothman, K. J. & Greenland, S. (eds.), *Modern Epidemiology (2nd edition).* Philadelphia: Lippincott Williams & Wilkins: 499– 510.

**TASK 1 – Define terms**

1. Write definitions or descriptions of the following terms in your own words. Use the two readings above to clarify your understanding. On page 112 of Chapter 6, Bonita et al., (2006) you will find explanations of various types of screening.
2. Illustrate each definition with an example that is relevant to your own work or experience. This is an extremely useful way of learning to apply new terminology.

screening targeted screening case-finding sensitivity

lead time

mass screening multiple screening diagnostic test specificity

detectable preclinical phase

#### FEEDBACK

Here are some definitions of the key terms:

|  |  |
| --- | --- |
| *Mass screening* | screening the whole population |
| *Multiple screening* | using a variety of screening tests simultaneously |
| *Targeted screening* | screening groups with known exposure to risk |
| *Case-finding or opportunistic screening* | screening that is restricted to patients who visit a clinic or doctor |
| *Sensitivity* | the proportion of truly ill people in the screened population, who are identified by the screening test |
| *Specificity* | the proportion of truly healthy people in the screened population, who are identified by the screening test |

## PRINCIPLES OF SCREENING

Not all screening activities are beneficial. On page 499 of Chapter 25 in Morrison (1998), the characteristics of diseases that make them suitable for screening are described. Essentially, the control of the disease must benefit from early detection, and, therefore, early treatment.

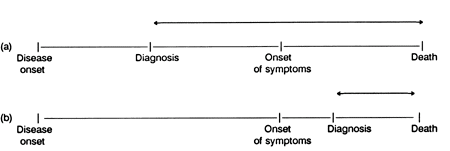
Assessing the benefits of early treatment is not always easy. One potential source of error is the phenomenon known as *lead time bias*.

#### Lead time bias

Suppose that we wish to explore the scope of reducing mortality from breast cancer by early diagnosis. One approach might be to compare the survival of patients whose tumours were detected at screening with that of women who only present once their disease has become symptomatic.

However, this could be misleading. Survival might be longer in the screened women not because early treatment is beneficial, but simply because their tumours are being diagnosed earlier in the natural history of their disease.

## Lead time bias



*Fig. 1: Lead time bias:*

*Group (a): Screening resulting in early diagnosis and treatment*

*Group (b): Diagnosis and treatment started after the onset of symptoms.*

With early screening and detection, the disease can be diagnosed earlier than without screening, resulting in lead time bias. In the figure above, it could appear that the patients in Group (a) survive for a longer time as they receive treatment for a longer time. However, this does not necessarily mean that the treatment has modified the course of the disease. That is, the time between disease onset and death could be the same for both Groups (a) and (b). This would mean that screening for the purpose of early detection and treatment is not beneficial.

Now compare two screening activities.

**TASK 2 – Assess two screening activities**

Some health systems routinely encourage women over the age of 35 to submit themselves to an annual gynaecological examination, during which a small scraping of cells (PAP smear) is taken from the cervix for testing. The aim is to detect any early signs of cervical cancer.

Another commonly promoted screening activity for women is regular breast self- examination for abnormalities that may represent the early signs of breast cancer.

Assume that breast cancer and cervical cancer rates are fairly high in the community where these activities are promoted. We know these cancers are a serious problem.

.../pto

1. What are the advantages and disadvantages of screening for cervical cancer?
2. What are the advantages and disadvantages of encouraging the breast self- examination as a screening activity?

#### FEEDBACK

1. The PAP smear requires: a visit to a clinic, the collection of the scraping by a health professional, the laboratory diagnostic process and the reporting of results to the health care worker and the patient on completion of the laboratory process. This is a multi-stage, complex and expensive process.

Even if the test itself is very accurate in identifying pre-cancerous conditions, screening every woman over 35 in the country on an annual basis is a massive and costly undertaking. To make such a screening programme worthwhile, you would need to be convinced that you are detecting large numbers of potentially fatal cases or reducing expensive treatments for cervical cancer. You would also want to be sure that you were doing this for the group of women most at risk of getting cervical cancer. For this reason, 35 years of age might be a bit young as a starting point for screening.

1. Self-examination of the breast is a very different process. The main costs are the education and advertising process required to communicate the information and motivation that women need to effectively screen themselves for the early signs of breast cancer. This is clearly a lot less expensive as it does not require complex medical or laboratory procedures. These will only become necessary if the women detect an abnormality during their regular screening. A much smaller number of women would therefore require more intensive investigation. The clinical infrastructure required to service this smaller group of women is much less than the annual PAP smear screening system. Another concern is the ability of the breast self-examination to consistently pick up pre-cancerous changes.

Therefore, it is important to evaluate the advantages and disadvantages of proposed screening activities before implementing them.

**TASK 3 – Assess a screening programme using a set of criteria**

To assess whether to institute a screening programme, various textbooks and researchers, including the World Health Organisation (Wilson & Junger, 1968), have recommended the criteria listed in the table below.

1. Re-assess the two examples in Task 2 using these criteria to determine which screening activity should be promoted.

## Criteria for instituting a screening programme

|  |  |
| --- | --- |
| Disease | Must be serious.  Shows high prevalence of pre-clinical stage. Natural history is well understood.  Long period between first signs and overt disease. |
| Diagnostic test | Sensitive and specific. Simple and cheap.  Safe, acceptable (to the people on whom it is used). Reliable. |
| Diagnosis and treatment | Facilities are adequate.  Effective, acceptable and safe treatment available. |

In the next task you will look at the accuracy of screening tests.

## INTERPRET SCREENING RESULTS

Even if prognosis is improved by early treatment, screening is only worthwhile if a satisfactory diagnostic test is available. The test must detect cases in sufficient numbers and at acceptable cost, and it must not carry side effects that outweigh the benefits of screening.

#### Sensitivity and specificity

Review the definitions of *sensitivity* and *specificity* on page 112 of Bonita et al. (2006). Another way of understanding *sensitivity* is to define it as the ability of the test to detect true positives. In a test with **high sensitivity**, a high proportion of the people that are classified as “positive” by the test must in truth have the disease or risk factor. This may also mean that the test is likely to produce many “false positives” – i.e. classify people as having the disease when in truth they do not - because its ability to detect those who do not have the disease or risk factor is weak.

Similarly, *specificity* can be understood as the test’s ability to detect true negatives. In a test with **high specificity**, a high proportion of the people that are classified as “negative” by the test truly do not have the disease or risk factor. This may also mean that the test is likely to produce many “false negatives” – i.e. classify people as not having the disease when in truth they do - because its ability to detect those who have the disease or risk factor is weak.

The second last paragraph on page 112 of Bonita et al. (2006) explains the trade-off between sensitivity and specificity.

**TASK 4 – Assess the validity of screening**

1. Under what circumstances would you prefer a screening test with high sensitivity to one with high specificity? Provide examples of diseases where this would be applicable.
2. Under what circumstances would you prefer a screening test with high specificity to one with high sensitivity? Provide examples of diseases where this would be applicable.

#### FEEDBACK

* 1. Tests with high sensitivity are selected for diseases where survival is critically dependent on early or immediate detection, e.g. maternal health, illness in children and particularly in newborns.
  2. Tests with high specificity are selected for diseases and conditions where the diagnosis is irreversible and could lead to stigmatization, e.g. mental health conditions (psychopathologies). High specificity is also used for diseases where follow-ups require a battery of expensive tests to confirm the diagnosis, e.g. cancers.

#### Predictive value

In addition to sensitivity and specificity, the performance of a test is measured by its *predictive value*. The predictive value of a positive result is the probability that a person who reacts positively to the test actually has the disease. Predictive value varies with the prevalence of disease in the population where the test is applied. If the prevalence is low then there are more false positive results than true positives, and predictive value falls. In this case, we refer to the screening test as having a low *yield*.

Probably the most well-known and propagated screening activity in sub-Saharan Africa today, is the HIV-test, or better known Voluntary Counselling and Testing (VCT) for HIV. The task in the following section requires you to reflect critically on testing for HIV, taking in consideration its social context.

## REFLECT CRITICALLY ON THE PRACTICE OF SCREENING

HIV testing is promoted widely as the first step towards receiving HIV care and treatment, and to provide impetus to HIV prevention efforts. The benefits of testing for HIV are numerous. Women who are pregnant and who test positive, can enter PMTCT programmes, where they would receive counselling on how to prevent their babies from getting infected. Testing positive in a Public Health facility gives you access to HIV care and treatment. In addition, knowing your HIV status has been shown to effect positive behaviour change as well as to have psychological and social benefits.

**TASK 5 – Assess the advantages and disadvantages of screening for HIV**

1. Comment on the various tests used for HIV screening. What is the accuracy (sensitivity and specificity) of these tests?
2. Are there any disadvantages to screening for HIV? (Hint: consider ethics)
3. Can testing for HIV ever be “bad” for anyone? Are there any circumstances where HIV testing should *not* be encouraged?
4. Given all the considerations above, what would you recommend in terms of testing for HIV in a Public Health service?

#### FEEDBACK

1. The most common test used in HIV screening is the anti-body test. This test seeks to detect the presence of HIV anti-bodies in the blood. The Elisa test for example is a test for HIV antibodies. Anti-body tests have a calculated sensitivity of 99.7% and specificity of 98.5%. With such high sensitivity, the chances of getting a false positive result is great; therefore a confirmatory test (normally the Western Blot) is always used to confirm positive result.

A viral test (such as the PCR) is more accurate and more expensive. It is generally used for screening infants born to HIV-positive mothers. This test seeks to detect genetic material of HIV.

1. When testing for HIV is done in the absence of treatment and care being made available to people testing positive for HIV, it is then considered unethical to screen.
2. In addition to testing being not advisable when there is no concomitant treatment or care available, testing mentally unstable patients such as diagnosed depressives and suicidal patients, is not advised. Consideration should also be taken for patients with drug dependencies. In all cases it is not a matter of denying such patients the HIV test, but rather delaying testing until adequate counselling and support services exist to allow these patients to safely deal with the shock of (potential) HIV seropositive status and to comply with an appropriate care or treatment programme.
3. Refer to the following websites for more information.

**READINGS**

<http://en.wikipedia.org/wiki/HIV_test>

<http://www.who.int/hiv/pub/vct/statement/en/index.html>

## SESSION SUMMARY

In this Study Session, you have explored the concept, purpose and methods of screening and reflected on the benefits of screening in the health context, taking in consideration the variations in accuracy of the tests as well as the social context.

The next Study Session looks at surveillance as a way of controlling communicable diseases. You will examine various methods of surveillance and assess a surveillance experience related to HIV/AIDS. You will also have the opportunity to critique the system and process for the notification of communicable diseases.



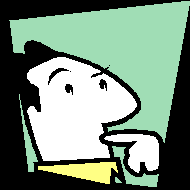
Screening does not diagnose people with a disease, but identifies those people who are likely to have the disease, in order to orientate them toward further investigations.

Criteria used to select a screening test are of three types: disease related, health-system related and test related.

**KEY POINTS IN UNIT 2 SESSION 4**

**Screening**

**Self-evaluation:** Try to imagine … what are the consequences of using a screening test with a low sensitivity or a low specificity?



# Unit 2 - Study Session 5 Surveillance

## Introduction

Surveillance is an essential part of the control of communicable disease, and is often achieved by means of the routine notification system. Any occurrence of cases that are clearly in excess of what would normally be expected can be easily and quickly identified. This allows a swift and effective response from the Public Health authorities.

The steps in disease surveillance have been clearly documented by the World Health Organisation (WHO) in field guides for the priority diseases of the Expanded Programme for Immunisation (EPI). In South Africa, these priority diseases are polio, neonatal tetanus and measles. What are the priority diseases in your country?

This Study Session looks at the process of monitoring disease in the community using different methods of surveillance.

## Contents

1. Learning outcomes of this session
2. Readings
3. Define new terms
4. Assess an HIV/AIDS surveillance experience
5. Critique the notification process
6. Session summary

## Timing of this session

This Study Session contains two readings and five tasks. It should take you about two hours to complete.

## LEARNING OUTCOMES OF THIS SESSION

|  |  |
| --- | --- |
| **By the end of this Study Session you should be able to:** | |
| ***Health Measurement Outcomes***   * Describe the role of surveillance, notification and special investigations. * Apply criteria for the introduction of surveillance programmes. * Critique the notification process in the country. | ***Academic Learning Outcomes***   * Define and apply new terminology. * Predict the likely result and impact of a variety of options. * Assess the impact of a surveillance programme. * Critique a national process. |

* 1. **READINGS**

There are two readings in this Study Session. You will be referred to them where relevant.

|  |  |
| --- | --- |
| **Author/s** | **Publication Details** |
| Vaughan, J. P. & Morrow, R. H. | (1989). Ch 5 – Reporting and Surveillance Systems. In *Manual of Epidemiology for District Health Management.* Geneva: WHO: 45–58. |
| Joubert, G., Ehrlich, R., Katzenellenbogen,  J. M. & Abdool Karim,  S. S. | (2007). Ch 13 – Routine health information systems and disease surveillance. In *Epidemiology: A Manual for South Africa.* 2nd Ed. Cape Town: Oxford University Press: 172–187. |

## DEFINE NEW TERMS

As with each of the other areas of health measurement, surveillance also has a number of new terms and concepts to describe different aspects of the health and disease monitoring systems currently in use.

The task that follows clarifies these terms.

**READING**

Vaughan, J. P. & Morrow, R. H. (1989). Ch 5 – Reporting and Surveillance Systems.

*Manual of Epidemiology for District Health Management.* Geneva: WHO: 45–58. Joubert, G., Ehrlich, R., Katzenellenbogen, J. M. & Abdool Karim, S. S. (2007). Ch 13 –

Routine health information systems and disease surveillance. In *Epidemiology: A Manual for South Africa*. 2nd Ed. Cape Town: Oxford University Press: 172–187.

**TASK 1 – Define terms**

1. Write definitions or descriptions of the following terms in your own words. Use the two readings above and the glossary in Chapter 14 of Vaughan & Morrow (1989) to clarify your understanding.
2. As in previous tasks of this sort, illustrate your definitions with an example.

notification notifiable disease probable case routine surveillance sentinel surveillance

outbreak investigation active surveillance

extended programme of immunisation (EPI) possible case

definite case community surveillance special searches surveys

passive surveillance

#### FEEDBACK

Here are some of the key terms:

|  |  |
| --- | --- |
| *Routine surveillance* | every case of a particular condition seen must be reported and counted |
| *Sentinel surveillance* | uses data from a few selected sites rather than data from all sites |

In the next section, we will explore surveillance activities in more detail.

## ASSESS AN HIV/AIDS SURVEILLANCE EXPERIENCE

Before you examine surveillance activities, it would be useful to clarify the purpose of such systems. There are a number of important uses of surveillance listed below.

A good surveillance system can enable you to:

* Report morbidity and mortality
* Document distribution and spread of diseases
* Establish long-term trends in disease occurrence
* Detect epidemics
* Identify high-risk groups
* Estimate the magnitude of a health problem
* Facilitate planning of control and prevention strategies
* Evaluate interventions
* Guide resource allocation in Public Health planning
* Set research priorities
* Provide information about the natural history of certain diseases

In the next task you will apply your understanding of surveillance methods to a specific example.

#### TASK 2 - Assess the sentinel surveillance system in Zimbabwe

HIV sentinel surveillance, or surveillance from a few specific sites, has been operational in Zimbabwe since 1990. There are more than 22 sites where such sentinel activity is taking place. These sites include rural areas, growth points, commercial farming areas, small mining areas, medium size urban centres and the three major urban centres in Zimbabwe.

At these sites, the population groups that have been under surveillance are pregnant women attending antenatal clinics (ANCs) and patients with sexually transmitted diseases (STD) attending outpatient facilities or STD clinics.

At prenatal care clinics, blood which is earmarked for other tests (e.g. syphilis in pregnancy) is also tested for HIV. A similar process occurs for STD patients. The tests are carried out on an unlinked and anonymous basis. The selection process of people to be included in these studies varies in different sites. Random, systematic and convenient sampling procedures have been used. Sample sizes have also varied from year to year at the same sites and for the same group under surveillance.

1. Why is there a need to have all the various sites indicated in this surveillance effort? (i.e. rural areas, growth points and so on)
2. Discuss the limitations of data derived from sentinel surveillance, given the process of patient selection.
3. What needs to be done to improve the quality of surveillance?

#### FEEDBACK

1. As sentinel surveillance does not collect data from the whole population, the sites selected must offer the best possible representation of the population at large. All parts of the society must be included. For this reason a wide variety of settings in which data is collected, is desirable.
2. This particular example relies on the testing of those who attend STD clinics and women attending ANC clinics. Since not everyone is tested, the sentinel system may miss important groups of people who may also be at risk of having HIV. You need to ask who attends these clinics and who does not. This sample includes those who are more sexually active (with more partners and a higher probability of STDs) and sexually active women of childbearing age who are now pregnant. All attend the public sector clinics. Men are under-represented. Less sexually active people, those without STDs, people who attend private medical clinics and women who are not pregnant, are all omitted. Such factors could have a substantial influence on the interpretation of the surveillance results, unless you know the HIV rates in these other groups.
3. If the system is extended to include these under-represented groups the surveillance data would be more reliable.

Probably the most well-known surveillance system is that dealing with the notification of infectious and other diseases. The tasks in the following section look at notifiable diseases and the valuable information that notification provides.

## CRITIQUE THE NOTIFICATION PROCESS

The routine notification of a finite list of infectious and non-communicable disease or health conditions is a central part of the surveillance system in most countries.

**TASK 3 - Identify some conditions on the notifiable diseases list**

1. What conditions are on the Notifiable Diseases List in South Africa / in your own country?

#### FEEDBACK

**Notifiable diseases in South Africa (2010)**

|  |  |
| --- | --- |
| * Acute flaccid paralysis * Anthrax * Brucellosis * Cholera * Congenital syphilis * Crimean-Congo haemorrhagic fever and other haemorrhagic fevers of Africa (Dengue Ebola, Lassa, Marburg and Rift Valley fevers) * Diphtheria * Food poisoning * Haemophilus influenza type B * Lead poisoning * Legionellosis * Leprosy * Malaria * Measles * Meningococcal infection * Paratyphoid fever * Plague | * Poisoning from agricultural or stock remedies * Poliomyelitis * Rabies (human) * Rheumatic fever * Tetanus * Tetanus neonatorum * Trachoma * Tuberculosis: primary, pulmonary, other respiratory organs, of meninges, intestine, peritoneum, miliary, bones and joints, genitor- uninary system and other organs. * Typhoid fever * Typhus fever (louse and rat flea borne) * Viral hepatitis A, B, non-A non-B and unspecified * Whooping cough (Bordetella pertussis) * Yellow fever |

(Department of Health 2010 – [www.doh.gov.za/docs/dns-f.html)](http://www.doh.gov.za/docs/dns-f.html))

**READING**

Vaughan, J. P. & Morrow, R. H. (1989). Ch 5 – Reporting and Surveillance Systems. In

*Manual of Epidemiology for District Health Management.* Geneva: WHO: 45–58.

**TASK 4 - Describe the notification process in your area**

1. Briefly describe the system for notification of diseases in your health facility and area. Explain how the notification system reaches a National level.
2. How frequent and how accessible are the reports on notifiable diseases?
3. What are the common faults in the notification system?
4. Use the checklist in Vaughan & Morrow (1989), pages 57–58, to evaluate this system.
5. Propose what could be done to improve the system.

#### FEEDBACK

Having compared what you noted down about the way notification works (or fails to work!) in your area of the health system, you should have been able to identify ways to improve the system using the checklist in Chapter 5 of Vaughan & Morrow (1989).

In South Africa, the notification of diseases follows this route:

|  |  |
| --- | --- |
| **National**  Epidemiology and Surveillance Unit | |
|  | |
|  |  |
| **Province**  Health Information/Epidemiology Unit | |
|  | |
|  |  |
| **District**  Health Information Manager | |
|  | |
| **Sub-District (Local Municipality)**  Health Information Manager | |
|  | |
| **Health Facility (Hospital, CHC or clinic)**  Patients – Health Workers – Laboratory | |

You are now asked to consider the topical issue of HIV in relation to the list of notifiable

diseases.

**TASK 5 - Consider the inclusion of HIV as a notifiable disease**

Including HIV on the list of notifiable diseases has been a contentious issue and continues to be hotly debated in different countries around the world. Just recently the South African Ministry of Health was on the verge of adding HIV to the list.

1. List some of the points for and against adding HIV to the list of notifiable diseases.
2. What would you have advised the Ministry of Health to do?

#### FEEDBACK

The debate on this matter is far from resolved. Early in 1999, the South African Ministry of Health (MOH) decided to add HIV to the list. They thought that this would give them a better idea of the size, shape and rate of growth of this important epidemic. It seems that they did not seriously consider other, more effective ways of getting this information.

If you check the criteria for instituting a screening programme in Study Session 4, you will notice that HIV does not satisfy the acceptability criteria.

The MOH proposal meant that any person who tested positive for HIV would, by law, have been identified by having their name and address recorded in the notifiable diseases records of their local health authority (LHA). These records are usually summarised and reported to the superintendents and management teams of all parts of the health service in the LHA area. Since there is still a profound social stigma and great potential for discrimination against people with HIV, this could have a disastrous impact on the lives of individuals. This also introduces an ethical argument around the compromise of doctor-patient confidentiality.

What in the end happened in the South African case? The instruction to notify cases of HIV met with an almost complete refusal by health workers to implement it. This non- compliance with the new rule ensured its failure and eventual withdrawal. In another instance, in France, the government established an anonymous HIV notification system. The system was made anonymous after pressure from HIV-activists.

## SESSION SUMMARY

In this Study Session, you have explored the concept, purpose and methods of surveillance and examined the routine recording of notifiable diseases as part of the surveillance mechanism.



Before starting a surveillance programme, one needs to ask “what do we want to monitor or measure, why and how?”

Surveillance can be active, where one sets out to find data (such as in research or in health intervention assessment) or passive, where one waits for data to arise through a system already in place (such as routine surveillance, notification or sentinel surveillance systems).

**KEY POINTS IN UNIT 2 SESSION 5**

**Surveillance**