**1**

Introduction to

Measuring Health and

Disease

**Introduction**

This unit serves to introduce you to quantitative measurement in the health sphere and then proceeds to introduce you to the disciplines of epidemiology and statistics and thereafter introduces basic concepts in causality. A deliberately broad picture is provided in this introductory unit to allow you an overview of the simplicity of measuring health and disease states and yet initiate you into the complexity which this measurement involves.

**Contents**

Session 1: Introduction to quantitative measurement in public health

Session2: Introduction to Epidemiology

Session 3: Introduction to Statistics

Session4: Introduction to Causality in Epidemiology

**Timing**

This unit has several readings and tasks and it should therefore take you 8 hours to complete.

**Learning outcomes**

* Appreciate the value of quantitative measurement in public health
* Define and describe epidemiology
* Define and describe statistics

Explore the ramifications of the concept of causality

**Readings**

This module guide, as with all the module guides provided by the UWC School of Public Health, contains explanatory text which is bolstered by an array of readings. The explanatory text usually introduces the topic/s, provides relevant background information, attempts to present a contextualised overview of the topic/s and links it/them to what you have already covered in the modules as well as what you will be covering later on in the module. The readings then provide the details or the meat of the topic/s and hence it is crucial for you to work through them carefully in order to fully comprehend the topic/s they cover. The readings that will be used in this unit are listed below. They are also displayed at the points in the unit where it is indicated that it is appropriate to proceed to work through them.

Silman, A. J. and MacFarlane, G. J. (1995) *Epidemiological studies: a practical guide*. Ch 1. Scope of epidemiological enquiry and overview of main problem areas. Cambridge: Cambridge University Press. Pages 3–7.

Detels, R. (2004). Ch 6.1- Epidemiology: the foundation of public health. In Detels, R. McEwan, R. Beaglehole & H. Tanaka (eds.). *Oxford textbook of public health*. Oxford: Oxford University Press. Pages 485-488.

Hennekens, C. H. and Buring, J. E. (1987) *Epidemiology in Medicine*. Boston: Lippincott Williams & Wilkins.

Bless, C. & Higson-Smith, C. (2004) Fundamentals of Social Research Methods: An African Perspective. Cape Town: Juta. Pages 2 – 7; 21.

Bruce, Pope & Stanistreet ( 2009) Quantitative methods for health research: a practical interactive guide to epidemiology and statistics. Ch 1. The Different Methods of Acquiring Knowledge. Chichester: John Wiley & Sons

Rothman, K.S. & Greenland, S. (1998). *Modern Epidemiology*. Ch 2. Causation and causal inference. Philadelphia. Lippincott, Williams and Wilkins. Pages 7 – 17.

Altman, D. (1991). *Practical statistics for medical research.* London: Chapman Hall. Pages 10-18.

Motulsky, H. (2010). *[Intuitive Biostatistics: A Nonmathematical Guide to Statistical Thinking](http://www.amazon.com/Intuitive-Biostatistics-Nonmathematical-Statistical-Thinking/dp/0199946647/ref=dp_ob_title_bk)* [(3rd edition](http://www.amazon.com/Intuitive-Biostatistics-Nonmathematical-Statistical-Thinking/dp/0199946647/ref=dp_ob_title_bk)). USA: Oxford University Press. Pages 17-20.

**Tasks**

With every topic covered there are several tasks which are designed to assess whether you have assimilated the knowledge required to successfully complete the task and whether you can apply that knowledge to solve the problems posed in the task. It is therefore crucial that you attempt every task. Once you have attempted it you should then look at the feedback provided to check to what extent you have been able to successfully apply the knowledge learnt to solve the problem. If there are parts of the problem that you did not successfully attend to, then you will be able to see where you were deficient and you should go back over those sections of the module text and readers that covers those areas, in order to solidify your understanding. Feedback on the tasks is provided at the end of each study session.

To obtain maximum benefit from the module guide and readers you should obviously work through all of the tasks before looking at the feedback. The tasks will also be extremely helpful to you when answering questions set in the assignments, as the questions set in the assignments will have the same structure and format as the tasks.

Unit 1 – Session 1: Quantitative Measurement in Public Health

As the name suggests this course is about both measuring levels of disease amongst populations and measuring their levels of health. Measuring diseases is relatively simple provided that a diagnosis of a specific disease can be made, as one would then simply group and count the number of people with that disease and then describe what things those people have in common (besides the disease).

Measuring health is a bit more complicated. Health has famously been described by the World health Organisation (WHO) as not merely the absence of disease but rather as being in a state of complete mental, physical and social well-being. Clearly this idealistic state described by the WHO (and others would add spiritual well-being) is unattainable, except possibly for a very short time, however it did shift the focus of the health care industry onto attaining and maintaining an acceptable level of health. Gradually this acceptable level of health became almost intuitively defined as a level of health which allows an individual to function adequately within their physical and social environment.

Following this, various measurement scales could be developed, based on answers to a series of questions, to pragmatically determine levels of health. In conjunction with this measurement of the presence of factors that foster disease, could be developed to assess to what extent groups are healthy or not and which factors are likely to shift the balance from health to disease. These factors could be general socio-economic conditions such as poor housing, lack of sanitation, inadequate water supplies, insufficient food, air pollution and unsafe working conditions which diminish health and predispose to several illnesses; or specific factors which cause illness such as tobacco inhalation, alcohol consumption, bacterial and viral infections and insecticide poisoning. Completing the picture would be measurements of what types of health care services are available to combat disease when present, prevent disease before it arises and promote and maintain health.

Throughout the history of measurement in the health sphere, there has been a focus on measuring disease. Who has the disease? How many people have the disease? What happens to them after they develop the disease? What could be done to contain the disease and how could one prevent the disease in future? This was the case even though for at least four thousand years people have to some extent known that the circumstances in which they live, affects their health. More than four thousand years ago relatively large cities with populations of more than 50 thousand people, such as Mohenjo-Daro in India, Harrapa in Pakistan, Kahun and Memphis in Egypt, had building regulations for dwellings, provided bulk water for the populace and had drainage systems. These actions suggest, but do not confirm, knowledge of the effect of the environment on the health and general well-being of the population.

Since at least the time of time of Hippocrates (2500 years ago) it has been known that the environment in which one lives (Hippocrates describes the climate, soil, winds and water as affecting health), the amount of food one eats (eating to excess), the type of work one engaged in (type of pursuits) as well as one’s social habits (exercise and alcohol consumption) affect the type of disease one might develop. Despite this knowledge, measurement of these predictors of disease was seldom undertaken before the 19th century. Similarly, little effort was exerted on assessing those who remained free of disease. Again this is understandable since if one was free of disease then the interest of healers’ in one’s status would have been negligible.

During the 19th century as efforts were made to prevent infectious diseases by cleaning the environment, removing sewerage safely, enforcing housing regulations and improving water supplies, so an interest in measuring both levels of cleanliness and water supply developed. Regular measurements of these factors affecting the likelihood of disease then ensued, becoming common place in Europe. These measures were subsequently linked to already existing death records, allowing a more specific link between living circumstances and likelihood of death to be made. William Farr from 1839 onwards provided annual analytical reports of deaths in England. Soon after, these sanitary measures could even be linked to specific diseases, as in the famous example of John Snow linking water supplies to an epidemic of cholera in 1854. He found that people living in houses that received water contaminated by sewerage were much more likely to develop cholera than those who lived in houses that received water uncontaminated by sewerage. He therefore concluded that something in the sewerage-contaminated water caused cholera.

With the advent of health prevention interventions such as vaccination against smallpox using live cowpox virus (introduced by Edward Jenner in 1796) and advocating hand washing to prevent transmission of infections (introduced by Ignaz Semelweis in 1847 specifically to reduce puerperal fever), the measurement of the effects of prevention interventions became prominent. This concerted measurement of the effectiveness of prevention programmes continued into the era of the global immunisation programmes in the twentieth century, especially smallpox vaccination (smallpox was eradicated globally in 1977), measles vaccination and poliomyelitis vaccination.

In the mid twentieth century measuring the potential causes of non-communicable disease became prominent as did measuring the effectiveness and efficiency of health care services provision. After the Alma Ata convention in 1978 where the principle of comprehensive primary health care was endorsed by all countries, health promotion activities increased and the measurement of efforts at health promotion such as growth monitoring of children became common place. The natural and logical extension then was to commence measuring the health of people using various indices. As expenditure on health care services rose in the 1970s, measuring the costs of health services and the inequitable spread of health services amongst population groups became prominent.

This course is entitled *Measuring health and disease,* as it hopes to cover all the basic techniques for the measuring of:

* Disease
* Immediate factors causing disease
* Underlying factors causing disease (factors that cause the immediate factors)
* Factors preventing disease
* Levels of health
* Factors promoting good health
* Health service provision coverage, acceptability, effectiveness, efficiency and equity

We hope to accomplish this using the measuring disciplines of epidemiology and biostatistics.

In the second study session you will be introduced to the measuring science of epidemiology and in the third study session to the measuring science of statistics. The fourth study session then covers the application of the concept of causality to epidemiology.

Unit 1 – Session 2: Introduction to

Epidemiology

Someone once said, “I hate definitions. They tell you nothing about the things they are defining”. The dictionary of epidemiology by John Last (2001) defines epidemiology as “The study of the distribution and determinants of health related states or events in specified populations and the application of this study to the control of health problems”.

To someone just starting to learn about epidemiology this definition is likely to be as useful to them as clothes are to a bear. However to someone who has already delved into epidemiology and unlocked some of its secrets, this definition would be a succinct explanation of the science of epidemiology. In this introductory session we hope to cover all the underlying assumptions in epidemiology as well as provide a broad overview of the scope of the discipline. In so doing we hope that by the end of this session Last’s definition of epidemiology will become crystal clear.

We will begin with a reading by Silman and Macfarlane (2002). It provides a short description of their version of the essential tasks of epidemiology and thereby attempts to inform readers of the scope of epidemiology. Go through this reading now and then attempt the task below.

**Reading**

Silman, A. J. and MacFarlane, G. J. (1995) *Epidemiological studies: a practical guide*. Ch 1. Scope of epidemiological enquiry and overview of main problem areas. Cambridge: Cambridge University Press. Pages 3–7.

**Task 1**

*What aspect of epidemiology do Silman and Macfarlane focus on?*

*Why do you think that they focus on this aspect of epidemiology?*

*Note that you will obtain the most value from this and all the other tasks in this module if you first attempt to answer it before looking at the feedback. Hence first jot down your thoughts on these questions before you proceed to look at the feedback provided.*

*Feedback is provided at the end of the study session.*

We now move on to a reading by Detels, which provides a brief history of epidemiology and then attempts to define epidemiology and finally describes the uses of epidemiology.

**Reading**

Detels, R. (2004). Ch 6.1- Epidemiology: the foundation of public health. In Detels, R. McEwan, R. Beaglehole & H. Tanaka (eds.). *Oxford textbook of public health*. Oxford: Oxford University Press. Pages 485-488.

**Task 2**

*Above we have described epidemiology as a science, yet Detels claims that it is more of an art than a science. Why do you think Detels claims it is an art?*

*Detels describes various uses for epidemiology. For each of these uses of epidemiology he provides one or more examples. Attempt to provide your own example for each of Detels uses of epidemiology.*

*Feedback is provided at the end of the study session.*

In order to better understand the current state of epidemiology, its role today and how it evolved it is useful to look in more detail at the history of epidemiology. Hennekens and Buring (1987) provide an overview of the history of epidemiology and the development of modern epidemiology from the 1950s onwards.

Go through this reading and then attempt the task below.

**Reading**

Hennekens, C. H. and Buring, J. E. (1987) *Epidemiology in Medicine*. Boston: Lippincott Williams & Wilkins.

**Task 3**

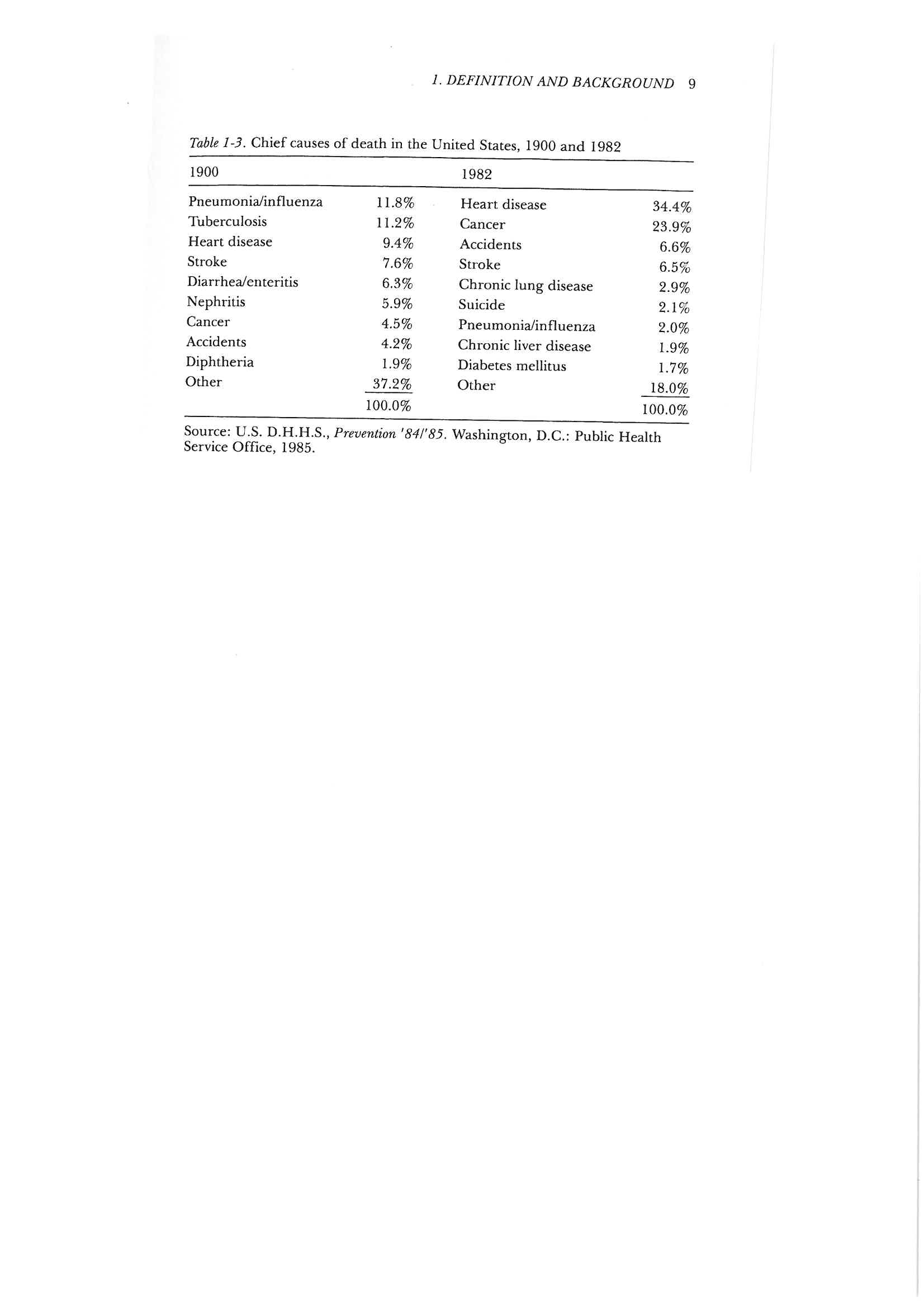
*Table 1.3 in the reading by Hennekens and Buring shows the top 10 causes of death in the United States of America (USA) in 1900 and 1982.*

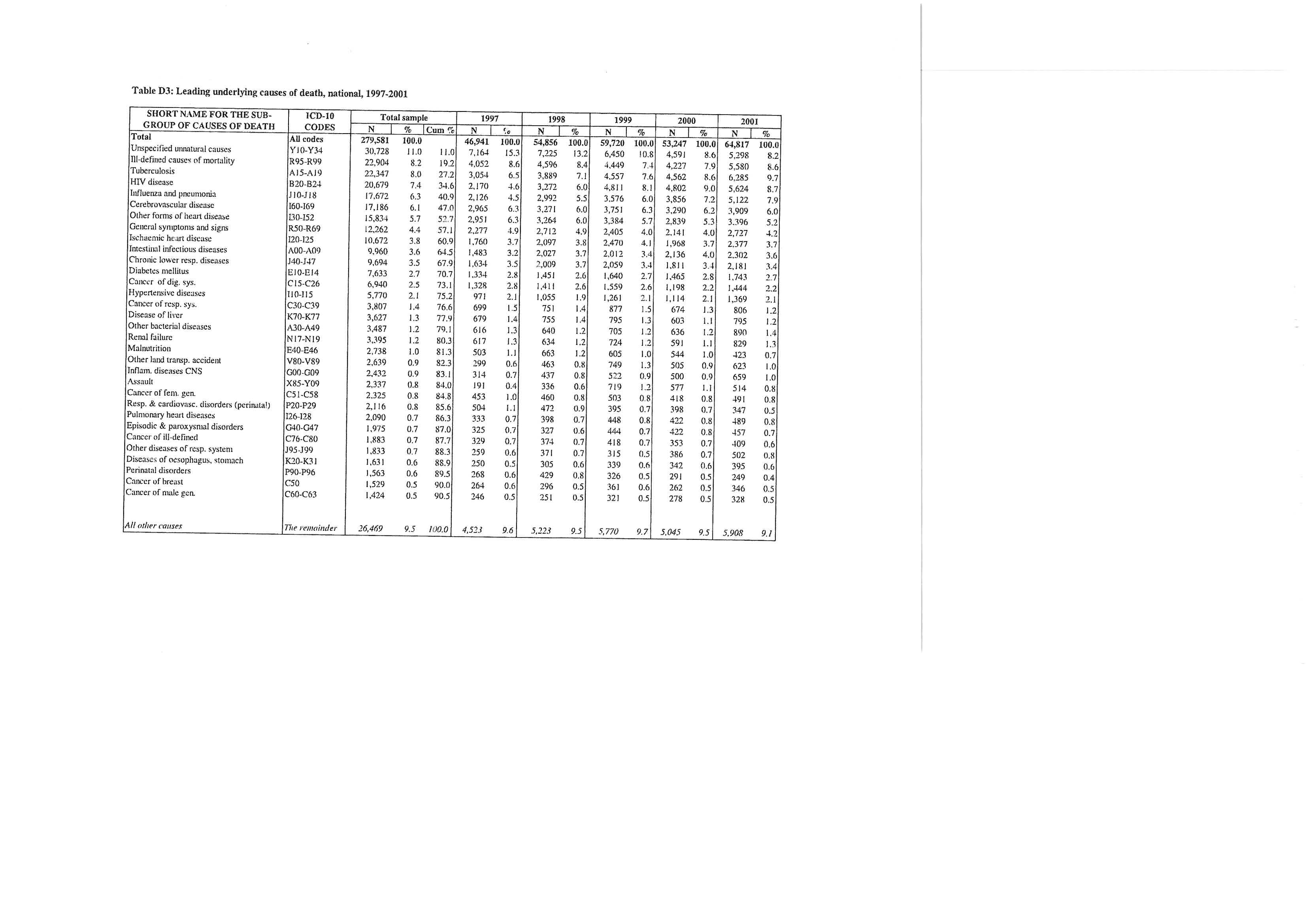
1. *How have the causes of death changed in those 82 years?*
2. *Why do you think they changed?*

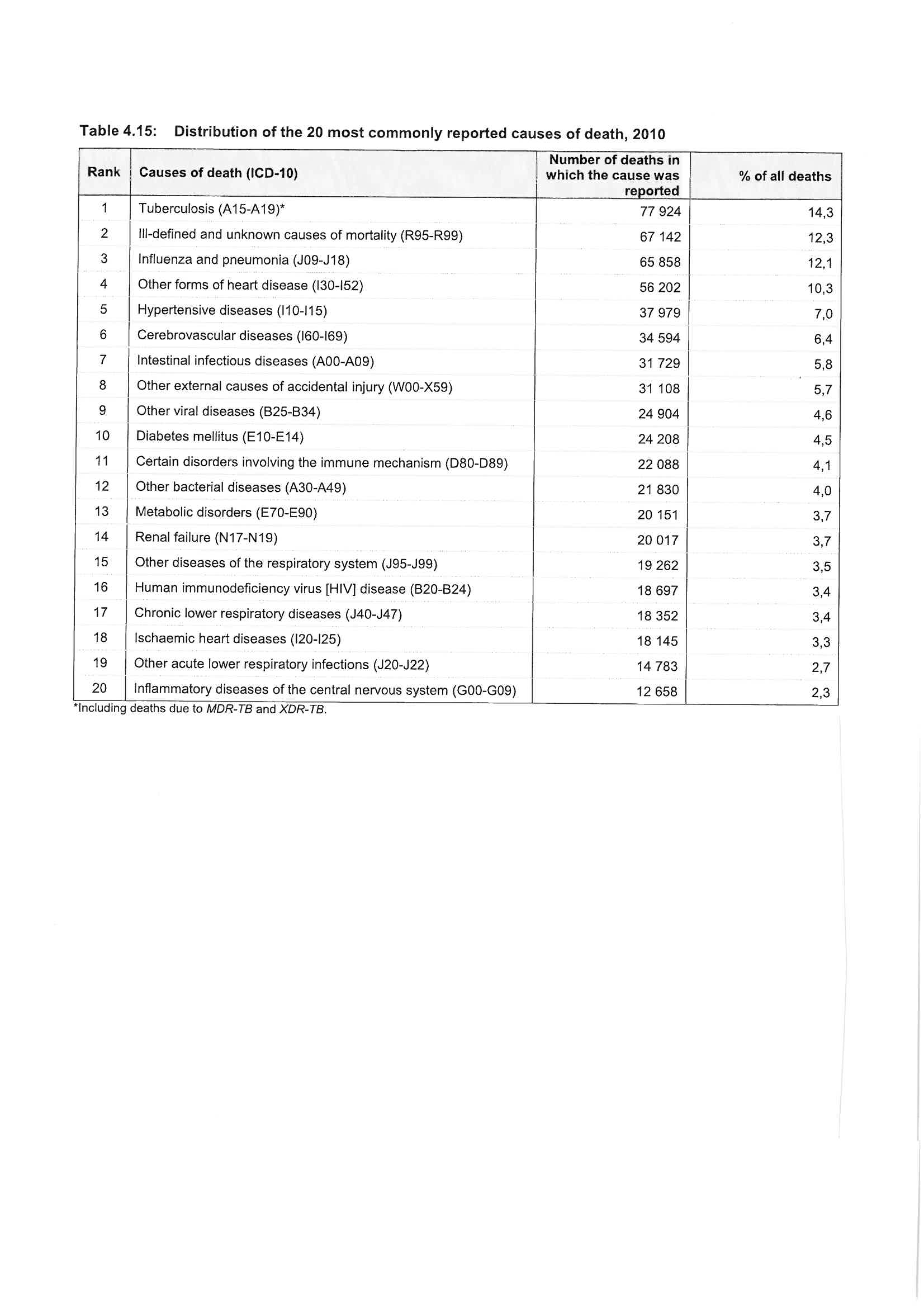
*Now look at the table (D3) below which shows the top causes of death in South Africa between 1997 and 2001.*

1. *What are the similarities and differences between the South African (SA) top ten causes of death and the USA 1900 causes of death?*
2. *What do you think is the reason for the pattern in the South African causes of death?*
3. *Then look at the top twenty causes of death in South Africa in 2010 shown in table 4.15 below.*
4. *How are these different from or similar to the SA causes of death between 1997 and 2001 and the USA causes of death in 1900, and why might it show this particular pattern?*
5. *Now finally compare the SA causes of death between 1997/2001and those in 2010 between USA causes of death in1982.*
6. *As before, no peeking at the feedback before you have provided your own answer to the task.*

*Feedback is provided at the end of the study session.*







In providing the feedback to the task above we made several assumptions regarding how things are linked through cause and effect. This requires that one has some prior knowledge and then uses that knowledge to interpret information. It also assumes particular ways of thinking which in turn are related to the manner in which knowledge is generated, transmitted and then criticised or accepted. These precursors to knowledge acquisition and ultimately the manner in which we use the knowledge, is based on the type or types of knowledge systems we were inducted into and our level of belief and trust in those systems.

The reading by Bess and Higson-Smith provides an introduction to various types of knowledge systems and in particular elaborates on the scientific method.

As you probably expected, you should now go through this reading and then attempt the task below.

**Reading**

Bess, C. & Higson-Smith, C. (2004) Fundamentals of Social Research Methods: An African Perspective. Cape Town: Juta. Pages 2 – 7; 21.

**Task 4**

*In interpreting the causes of death in the USA and SA, what type/s of knowledge systems did we use?*

*What typical assumptions did we make when we used these systems to explain the different and changing causes of deaths?*

*What are the dangers of interpreting the information on causes of death in the manner in which we did?*

*Would those dangers completely invalidate our interpretation of the causes of deaths?*

*Feedback is provided at the end of the study session.*

We have now introduced positivism as a philosophy of knowledge and since it underpins the foundations of epidemiology and statistics, it is useful to delve further into the basic theoretical underpinnings of positivism. In understanding positivism it helps to compare and contrast it to other philosophies of knowledge to better understand its assumptions and indeed its strengths and limitations. The major rival to positivism is interpretivism which views the world rather differently. Go through the reading by Bruce, Pope and Stanistreet (2009) which provides an overview of the positivist and interpretivist philosophies of knowledge. Thereafter attempt task 5.

**Reading**

Bruce, Pope & Stanistreet (2009) Quantitative methods for health research: a practical interactive guide to epidemiology and statistics. Ch 1. The Different Methods of Acquiring Knowledge. Chichester: John Wiley & Sons

**Task 5**

Why is the positivist philosophical approach most suited to the science of epidemiology?

*Feedback is provided at the end of the study session.*

**Feedback on Task 1**

*In this outline of epidemiology Silman and Macfarlane focus on the measurement of disease. They are very thorough about this and carefully explain how epidemiology can describe and assess all stages of the disease process in people. They probably focus on this as epidemiology arose from assessing how often diseases occur and trying to determine what causes particular diseases to occur amongst particular individuals.*

*Two crucial aspects of epidemiology which Silman and Macfarlane do not cover at all is that firstly epidemiology is as much concerned about health as it is about disease, and secondly that epidemiology does not in fact focus on individuals but instead assesses health and disease amongst groups of people. In the chapter by Detels (2002) in the Oxford textbook this theme is taken up and elaborated on. As before, go through this reading first and then attempt the task below.*

**Feedback on Task 2**

*Detels feels that epidemiology is sometimes an art as one needs artistic creativity to think of linking a particular variable to a particular outcome and identifying which population best suits the study of which particular epidemiological research questions. Also epidemiology is pragmatic and opportunistic, since it is carried out on humans, and therefore many epidemiological studies, if carried out in the best theoretical manner, would be impractical and unethical. This ethical and practical dilemma could then be circumvented by conducting the research on particular populations, since amongst a particular population, given their unique circumstances, the research would be practically possible and ethical. Hence in many instances when conducting epidemiological research one has to be opportunistic and think about what is possible in the various circumstances that obtain in various locations and points in time.*

*Note: You can email your list of examples on the uses of epidemiology to us and we will provide you with individual feedback on whether it is an appropriate fit for the use it is listed under.*

***Feedback on Task 3***

*A*

*In the USA in 1900 most of the deaths were due to infectious diseases (pneumonia, tuberculosis, diarrhoea, nephritis, diptheria), followed by non-communicable diseases (heart disease, stroke, cancer) with accidents low down in the top ten. By 1982 the proportion of people dying due to infectious diseases had decreased dramatically and non-communicable diseases are now overwhelmingly the main causes of death. Accidents have now become more prominent than infections and chronic lung disease makes its debut in the top ten causes of death. Note however that lung infections still remain in the top ten, although they are now much lower down in the top ten than before.*

*B*

*The pattern of causes of death probably changed as conditions changed in the USA. Improvements in nutrition, housing, sanitation, water provision and the arrival of antibiotics as an effective treatment for infections, all combined to dramatically reduce deaths from infections. It is likely that smoking tobacco, eating a diet higher in starch, fat and salt, decreasing physical activity levels, increased alcohol consumption and living to an older age, has resulted in non-communicable diseases becoming overwhelmingly the main causes of death by 1982. Accidents as a cause of death probably increased due to the massive rise in motor vehicle use over the time period. This was probably moderated by the improvement in safety conditions in the workplace resulting in less workplace accidents. Note that suicide has now crept into the top ten causes of death reflecting greater levels of mental illness which in turn is linked to changes in social conditions.*

*C*

*The causes of death in SA from 1997 to 2001 have several similarities with the causes of death in the USA in 1900. In SA 1997-2001 infectious diseases (tuberculosis, pneumonia, HIV, intestinal infections [diarrhoea]) are overwhelmingly the most common cause of death, as in the USA in 1900. In both countries in the respective time periods non-communicable diseases (cerebrovascular disease, ischaemic heart disease, other forms of heart disease, diabetes, cancer) and accidents/homicide (unspecified unnatural causes of death, other land transport accidents, assault) follow as the next biggest causes of death, but at much lower levels than infections.*

*The SA top ten causes of death differ from those in the USA in that an infection prominent in SA in 1997-2001, namely HIV, was not present in 1900 USA. This new infection had now surfaced and rapidly climbed the ranks of causes of death, becoming the third commonest infectious cause of death. The arrival of an entirely new (previously unknown) cause of death is typical of the rapid spread of an infectious disease, which has a high probability of a lethal outcome.*

*The other major difference between the SA 1997-2001 causes of death and the USA 1900 causes of death, is that in SA non-communicable causes although being a clear second to infections, are at higher levels than in USA 1900. Similarly accidents/homicide are a much more prominent cause of death in SA 1997-2001 than it was in the USA in 1900. Accidents as a high cause of death in SA probably reflects both motor vehicle accidents concomitant with the rise of motor vehicle usage coupled with the lack of safe public transport, and workplace accidents, especially in the mining, fishing and farming sectors, which remain very dangerous occupations for low rung workers.*

*D*

*This pattern of a large proportion of deaths due to infections and a rising level of non-communicable diseases and accidents/homicides in SA, is typical of a situation where there are diverse groups in a country, prone to different diseases due to their different living conditions. This pattern therefore reflects a situation where some groups are prone to all causes of death (infectious, non-communicable diseases and accidents/homicides) while other groups are only prone to some of them (e.g. non-communicable diseases). Or alternatively some groups of people are prone to some diseases (e.g. infections) and others are prone to other diseases (e.g. non-communicable diseases). Indeed the former situation prevails in SA , which was a politically, economically and racially oppressed country where the vast majority of the oppressor group lived in good socio-economic conditions. Due to good nutrition, housing, sanitation, water provision and health services (including antibiotics) infectious diseases were rare as a cause of death amongst this group. However, similar to the pattern in the USA, non-communicable diseases increased amongst this group as a result of smoking tobacco, eating a diet higher in starch, fat and salt, decreasing physical activity levels, increasing alcohol consumption and living to an older age. On the other hand the majority of the oppressed group who were living in poor housing with minimal or no sanitation, accessing insufficient clean water, lacking adequate nutritious food and lacking access to good health services, remained prone to infectious diseases. However as the lifestyle factors linked to non-communicable diseases, such as “junk food” (high calorie, high salt) diets and inactivity, became more pervasive in South Africa, due partly to globalisation, those amongst the oppressed group who had escaped an infectious death, became prone to dying from non-communicable diseases. Due to high levels of violence, low levels of security, exploitative and dangerous conditions of work, increasing numbers of vehicles, poor road safety mechanisms, inadequate and dangerous public transport and high levels of social instability, the deaths due to accidents/homicide were present amongst both groups, but were much higher among the oppressed group of people.*

*The high level of HIV as a cause of death in SA 1997-2001 and its complete absence in USA 1900, could only be explained by either it being a disease unique to SA or as a disease unique to the time period i.e. a new disease. Indeed, as is well known HIV is a new disease which did not exist amongst humans in 1900, but had reached epidemic levels in SA (and the rest of the world) by 1997-2001.*

*E*

*The causes of death in SA in 2010 show an unchanged pattern from that between 1997 and 2001, reflecting that insufficient change in living conditions and health services had occurred during the intervening 9 -14 years, to modify the causes of death.*

*F*

*The USA 1982 causes of death are overwhelmingly due to non-communicable diseases with infections being a very low cause of death but with accidents and suicide levels rising. In SA in 1997-2001 and still the same in 2010, infections are still by far the largest cause of death, but non-communicable diseases are rising and although accidents/homicide are static, they are static at a high level. These differences reflect the differing conditions in the two countries. So as detailed above, in South Africa the conditions for high levels of infectious disease and accidents/homicide still remain (and a new infectious disease HIV has arisen to add to the infectious disease burden), but at the same conditions promoting non-communicable disease are becoming more prevalent.*

**Feedback on Task 4**

*We used the rationalistic and the empirical methods which when combined constitute the scientific method. We had prior knowledge of the common causes of the diseases which resulted in deaths and in which we were very confident and we then simply assumed that if those causes were present then they were likely to be the causes of the diseases in the situations presented to us in the USA and SA.*

*The danger of doing this is that we are generalising from the general to the specific and simply assuming that what applies generally should also apply in the specific context. There is no good reason to make such assumptions (although it seems to be a natural tendency of humans), as there could be other causes of the diseases in those particular contexts and/or the general causes of the diseases might for some reason not be applicable in those contexts, rendering our interpretation invalid.*

*There is a slight chance that it could completely invalidate our interpretation, however our interpretation is saved from complete invalidation because of its logical consistency, its use of prior empirical findings and the probabilistic nature of our interpretation. We used terms such as probably, likely and could be, all of which indicate a degree of uncertainty and hence leave open the option that there might be other varying interpretations of the same information.*

*The scientific method is underpinned by the philosophy of positivism. Positivism is not the only philosophy of knowledge, but it is argued that it is the one most applicable to epidemiology.*

**Feedback on Task 5**

*Feedback is provided in the reader itself, at the end of the reading.*

Unit 1 – Session 3: Introduction to Statistics

When conducting quantitative studies a lot of data is obtained and these data are usually expressed as numbers. That leaves the researcher with the problem of how to make sense of all those numbers. Statistics is a branch of mathematics that helps us to make sense of large sets of numbers. Hence a basic knowledge of statistics is essential for all quantitative researchers.

Specifically statistics helps us to:

1. **Sort** (organise), **check and clean** a large set of data
2. **Summarise** large volumes of data using graphs and data summary concepts, after cleaning/sorting it
3. **Analyse** the data to answer the research question, after summarising it
4. **Interpret** the data to arrive at conclusions about what it is showing us, after analysing it
5. Decide on the **sample size** required for a particular study, before the study commences
6. Comment on the **limitations** imposed on the study by the **sample size** and the **degree of uncertainty** shown in the analysis,and hence what cautions we should apply to our conclusions drawn from the study.

So if statistics can do all that, then let’s look at how it could do that. For example if we measured the weight of one thousand children aged 1 year or younger, we would have a large set of data comprised of the weight of each of those thousand children. This gives rise to the immediate problem of how one handles this large amount of data in order to make any sense of it. The first step would be to sort the data. We could do this by listing the weights from lowest to highest in a table and by plotting them on a graph. The next step then would be to check the data. If on looking at the table or graph we find that the lowest weight listed is 100 grams, while the 2 highest weights listed are 150 kilograms and 1200 kg, we would be quite sure that those figures are incorrect. We are sure it is incorrect as it is inconceivable that any child could weigh that little or that much. So we would need to go back and re-weigh those children to determine what their actual weights are and then replace the incorrect figures in the dataset with those corrected figures. If we are unable to go back and weigh the children again, then we would have to exclude those records from the dataset, as we know they are incorrect.

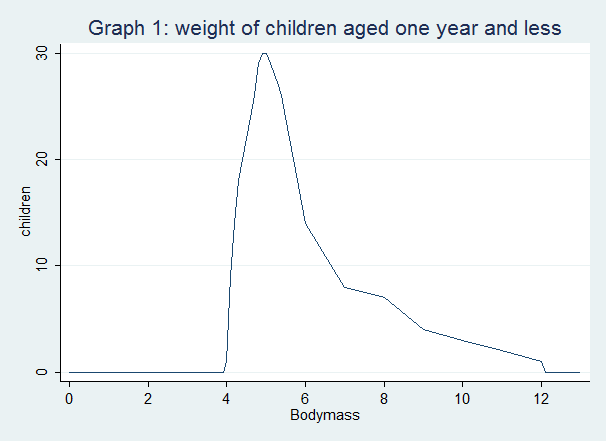
So having cleaned the dataset there would be a range of weights with some children having low weights and some children having high weights, and some children having in between weights. Perhaps we could start with what was the lowest weight and what was the highest weight. We would then know that all the other weights were between those two values, so we would know the spread or range of weights would fall between those 2 values. However we would not know if most of the other weights are close to the low weight, or if they are close to the high weight, or if they are evenly spread out between the low and high weights. So to assess this we would have to look at the spread of weights.

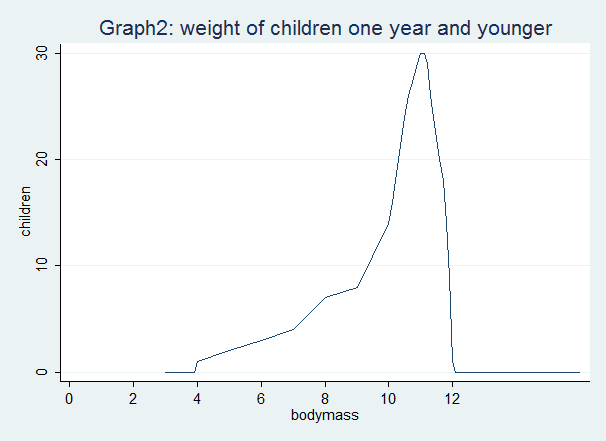
Have a look at the 2 graphs below. They are graphs of 2 groups of children aged 1 year or younger and both of them have a range from 4kg to 12kg. However in graph 1 most of the weights of the children are clustered between 4 and 6 kilograms, while in graph 2 most of the weights of the children are clustered between 10 and 12 kilograms. So having sorted the data from lowest to highest and then graphed it, we are able to see that although there is overlap between graph 1 and graph 2, in that both graphs have weights that range from 4 to 12 kg, the grouping of weights in graph 1 are very different from the grouping of weights in graph 2.

So how could we interpret and explain this difference?

If most of the children in both graphs are close to 1 year of age then it would be reasonable to assume that many of the children in graph 1 could be malnourished, as 1 year old children should weigh more than 4 to 6 kilograms.

Or if the ages of most of the children in both graphs are close to 3 months then many of the children in graph 2 could be obese, as 3 month old children should weigh less than 10 to 12 kilograms.



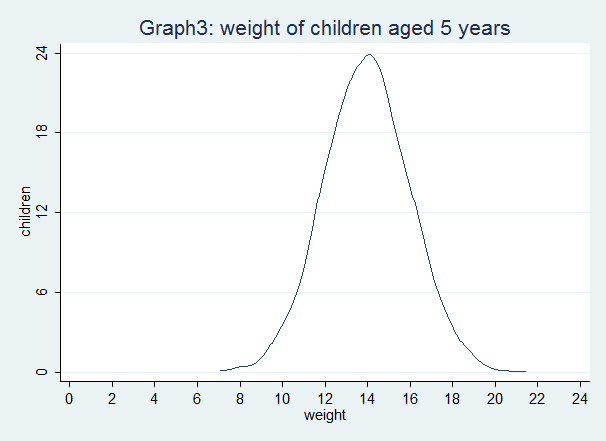


*Task 1*

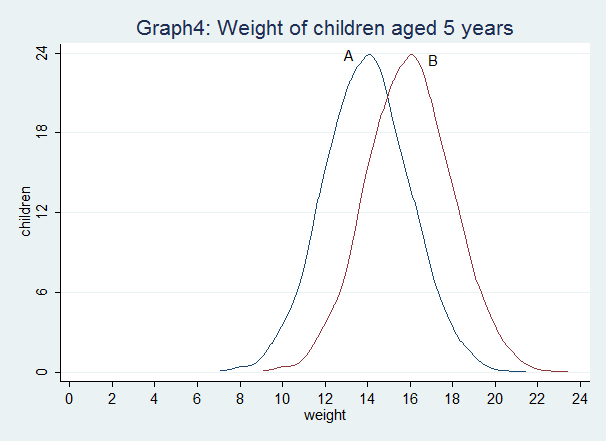
*What other explanation can you think of to make sense of the 2 graphs?*

*(Feedback is provided at the end of the study session)*

Graph 3 shows the range of weights of children who are all 5 years of age. What can we see from this graph? We can see that the range of weights is from 8kg to 20kg. We also notice that very few children have a weight close to 8kg. Similarly very few children have a weight close to 20kg. Most of the children’s weights are clustered around the 14kg mark. If you draw a dotted line from the 14kg weight mark on the x-axis (horizontal axis) to the point where the dotted line meets the line graph then we can see that the dotted line acts very similar to a mirror, with the weights to the left of the line forming a picture similar to the weights to the right of the line. This shows that although the weights range from 8kg to 20kg most of the children have a weight close to 14 kg, but some have a weight much lower than that extending to 8 kg (the lowest weight) and some have a weight much higher than that extending to 20 kg (the highest weight). We can thus deduce that the range of weights are symmetrical around a central point of 14 kg. Or stated in another way the average weight is 14kg, but the range of weights could be between 8 and 20 kg.



If we now wanted to compare this group of 5 year old children with another group of 5 year old children, we could do that by comparing the graphical pictures of their weights. We have done this in graph 4 below. So what can we determine by looking at the two groups of children shown in graph 4. Group A is our original group with a range of weights symmetrical around a central point of 14 kg, while group B is the other group with a range of weights symmetrical around a central point of 16 kg. So the groups are similar in that they are both consist of children who are 5 years of age and the weights in both groups are symmetrical around a centre point.



*Task 2*

*Describe what you see in graph 4.*

*Is there any difference in weights between groups A and B?*

*If there is a difference between the groups then describe the difference.*

*(Feedback is provided at the end of the study session)*

We have thus far briefly gone through how statistics assists us to **clean** (we uncovered and removed errors in weights), **sort** (we listed the weights from highest to lowest), **summarise** (we presented the data on a graph), **analyse** (we compared the two groups of children using graphs) and **interpret** (we described each group and determined if there were any differences and similarities between them ) quantitative data.

Now we will move on to consider how using a **sample** is a useful statistical technique. If one wanted to study some aspect of a large group of people, for example if one wanted to know what percentage of people in a city of 3 million people have arthritis, then one is immediately faced with the problem of having to check each of those 3 million people to see if they have arthritis or not. If one could instead just examine a small portion of the 3 million population, then one’s job would be easier. But would a small portion of the population give one the answer to the question, what is the percentage of people in the city who have arthritis? The answer is yes, if one selected that small portion or sample of the people in the city in such a way that they properly represent all the people of the city. Hence the first characteristic of a sample when conducting quantitative research is that it must properly represent the population that it is derived from. If it fulfills this criterion then the sample can be used to validly infer things about the population. The immediate benefit of a sample then is that it decreases the amount of measuring you would have to do. This in turn means that the amount of time, effort and money spent on the study would be reduced and hence using a sample makes a study much more efficient than measuring the entire population.

The reading by Motulsky provides a brief introduction on how to use data from samples to infer things about the population whom the sample represents. More details on using a sample to infer things on the population, as well as on the types of samples one can use and how to determine how big or small the sample should be and the limitations that using a sample imposes on a study, are provided in unit 2.

**Reading**

Motulsky, H. (2010). *[Intuitive Biostatistics: A Nonmathematical Guide to Statistical Thinking](http://www.amazon.com/Intuitive-Biostatistics-Nonmathematical-Statistical-Thinking/dp/0199946647/ref=dp_ob_title_bk)* [(3rd edition](http://www.amazon.com/Intuitive-Biostatistics-Nonmathematical-Statistical-Thinking/dp/0199946647/ref=dp_ob_title_bk)). USA: Oxford University Press. p 17-20.

We will now delve deeper into how one could **sort and summarise** quantitative data. At the beginning of this session we noted that when conducting quantitative studies a lot of data is obtained and these data are usually expressed as numbers. We further noted that this leaves us with the problem of how to make sense of all those numbers. We then proposed that we could initially sort and summarise them and then attempt to make greater sense of them. The most basic way to sort quantitative data is to decide what type of quantitative data they are. **Quantitative data** can be classified into **two basic types,** namely whether they are **categorical or numerical.** Categorical data refers to definite groups of things or people, where the things or persons being measured can be sorted into: 1 of only 2 possible categories (binary), more than 2 categories (nominal), and more than 2 groups that can be ordered (ordinal), for example, classification according to mild, moderate and severe.

Numerical data refers to data which takes only a number form, such as the age of a person or the weight of a person. Numerical data can be sorted into: a set of data where values are measurements that can assume any value within a specified range (continuous), for example, height, weight, age and a set of data where values are counts (0,1,2,3,...) of events (discrete), for example, the number of births in a week; the number of patients in a clinic. Numerical data can however also be formatted into groups and therefore numerical data can be converted into categorical data. So the actual ages of people could be grouped into categories such as child, adolescent, young adult, middle aged, elderly and old. If you group them into such categories then you must decide which ages fit into which category, so is middle age from 40 to 60, or is it from 30 to 50, or is it from 30 to 60. As you can see how you decide to group (categorise) numerical data would have a great effect on which data goes into which group and hence the analysis becomes problematic. Therefore it is best to retain numerical data for a population of people as numerical data and to sort and analyse it in that form. Note that this is what we did above for the weights of the children, i.e. we analysed the weights as numerical data. Note that the reverse is not true and so categorical data cannot be transformed into numerical data.

The reading by Altman provides a useful overview of the types of quantitative data you will encounter when conducting research into the health and ill health of people. The two most basic types he describes are categorical and numerical, but he also introduces other types of data such as ranks, ratios, rates, scores and scales. Go through this reading and then attempt task 3 below.

**Reading**

Altman, D. (1991). *Practical statistics for medical research.* London: Chapman Hall. p 10-18.

*Task 3*

*Various types of data are described in the table below.*

*Indicate what type of data each one is.*

*(Feedback is provided at the end of the study session)*

|  |  |
| --- | --- |
| *Description* | *Type of data* |
| *Suburbs in a city* |  |
| *Ages of children in a school* |  |
| *Highest grade passed at school amongst a group of mine workers* |  |
| *Weight (body mass) of new born children* |  |
| *Number of children immunised or not immunised against measles in a country* |  |
| *Number of children born to women in a province* |  |
| *Blood pressure levels of people in a district* |  |
| *Number of people in a district who have hypertension (high blood pressure)* |  |
| *Body mass index of a soccer players* |  |
| *Number of soccer players who are overweight* |  |
| *Level of pain amongst people with arthritis* |  |
| *Amount of time spent waiting to be seen at a clinic by people who attended the clinic* |  |
| *Religious affiliation of children at a school* |  |
| *Proportion of children immunised in each of 20 districts in a country* |  |
| *Blood sugar levels amongst a group of diabetic patients* |  |
| *Marital status of men who attend a clinic* |  |
| *Level of satisfaction with a service received by patients at a hospital* |  |

***Feedback on Task 1***

*It could be that the groups of children in graphs 1 and 2 have a different spread of ages. So although the ages of the children in both groups range from 1 day to 1 year, if the spread of ages were different in the 2 groups then there might be an alternative explanation for the differences in weight, other than that one group might be either malnourished or obese. So for instance if in graph 1, although the age range is from 1 day to 1 year, perhaps most of the children shown in that graph are aged between 2 and 5 months. However in graph 2 most of the children shown in that graph are aged between 10 months and 1 year. Then we could conclude that the different clustering observed in weights between graphs 1 and 2 are due to the different clustering of ages in the two groups. In graph 1 the ages are clustered around 2 to 5 months and hence this is reflected by the lower weights clustered around 4 to 6 kilograms. In graph 2 the ages are clustered around 10 months to 1 year and hence this is reflected by the higher weights clustered around 10 to 12 kilograms. So the difference in weights seen is entirely explained by the different spread in ages between the 2 groups, even though they both have the same age range.*

***Feedback on Task 2***

*It can be seen that both groups have symmetrically distributed weights around a central point or average. On average those in group A have a weight of 14 kilograms while those in group B have an average weight of 16 kilograms. The lowest weight in group A is 8 kilograms and the highest weight is 20 kilograms. The lowest weight in group B is 10 kilograms and the highest weight is 22 kilograms.*

*Yes, there is a difference in weights between groups A and B.*

*Children in group B are on average 2 kilograms heavier than children in group A. Several children in group A have a lower weight that the child with the lowest weight (10 kilograms) in group B. So the smallest child in group B is still bigger than several children in group A. Several children in group B have a higher weight than the child with the highest weight (20 kilograms) in group A. So the biggest child in group A is still smaller than several children in group B.*

***Feedback on Task 3***

*The types of data are indicated below.*

|  |  |
| --- | --- |
| ***Description*** | ***Type of data*** |
| *Suburbs in a city* | *Categorical: several categories* |
| *Ages of children in a school* | *Numerical: usually discrete if take to nearest completed year or month; but in reality continuous as can measure age to months, weeks, days, seconds, microseconds, etc.* |
| *Highest grade passed at school amongst a group of mine workers* | *Categorical: ordinal; categories range in order from lowest grade to highest grade; Ordinal because e.g. grade 8 is not twice as good as grade 4 and therefore it is not numerical (note that 8 years is twice as much as 4 years and hence age is numerical as noted above)* |
| *Weight (body mass) of new born children* | *Numerical: continuous* |
| *Number of children immunised or not immunised against measles in a country* | *Categorical: binary (2 categories; either immunised or not)* |
| *Number of children born to women in a province* | *Numerical: discrete; 4 children really are two times more than 2 children* |
| *Blood pressure levels of people in a district* | *Numerical: continuous* |
| *Number of people in a district who have hypertension (high blood pressure)* | *Categorical: binary (2 categories; either classified as hypertension based on some cut-off blood pressure level or some other criteria, or not classified as hypertension)* |
| *Body mass index of soccer players* | *Numerical: ratio; body mass index is obtained by dividing height by weight squared to obtain a ratio value* |
| *Number of soccer players who are overweight* | *Categorical: binary (overweight or not overweight;*  *or Categorical: ordinal (if use a range of categories of weight from e.g. very underweight to very overweight)* |
| *Level of pain amongst people with arthritis* | *Numerical: scale; if choose from a set range e.g. from 1 to 10 where 4 is twice as painful as 2 (note is it not discrete numerical as the range set is arbitrary and could e.g. be from 0 to 5 instead of from 0 to 10);*  *or Categorical: ordinal; if choose from set categories where the pain distance between different categories is not constant e.g. minimal pain, painful, severe pain, very severe pain* |
| *Amount of time spent waiting to be seen at a clinic by people who attended the clinic* | *Numerical: continuous* |
| *Religious affiliation of children at a school* | *Categorical: several categories* |
| *Proportion of children immunised in a country* | *Numerical: ratio; ratio or percentage is obtained by dividing the number of children immunised by the total number of children* |
| *Blood sugar levels amongst a group of diabetic patients* | *Numerical: continuous* |
| *Marital status of men who attend a clinic* | *Categorical: several categories (single, married monogamous, married polygamous, divorced, widowed, cohabiting)* |
| *Level of satisfaction with a service received by patients at a hospital* | *Categorical: ordinal if overall view given from very dissatisfied to very satisfied;*  *or Numerical: score if several aspects of satisfaction (e.g. courteousness of staff, quality of care received, cleanliness of facility, etc.) are given a score and then all are added up to give a composite score.* |

Unit 1 – Session 4: Introduction to Causality in Epidemiology

You will probably have gleaned by now (note the probabilistic nature of this statement) that a large proportion of the philosophy and science of epidemiology (note the ‘philosophy’ and ‘science’ and not the ‘art’) rests on a specific positivistic view of causation. The manner in which we think about and understand causation affects our understanding and application of epidemiology.

The evolution of public health, and indeed the evolution of humans as an industrial and technological society, has been linked to the concepts of causality and the scientific method. Causality is the understanding that one set of events leads to another. The second set of events is caused by the first and therefore if the first set of events happens, one can predict that the second set of events will follow. The first set of events is referred to as a cause and the subsequent set of events as an effect. So for instance if a student does not study a particular set of learning materials, one could predict that the student will fail an examination that is based on the material the student should have studied (cause - lack of studying; effect - failing the examination). If one set of events depends on another, then it is possible to prevent that set of events from occurring by avoiding or counteracting the preceding event (cause) that leads up to and causes the set of events (effect). So for instance failing an examination can be prevented by studying and being familiar with the learning materials on which the examination is based.

In the same way unpleasant events such as illness, injury and death could be prevented or minimised if one knew what caused the injury, illness or death. Illness or injury could also be treated if one knew what could cure or control it. Public health improvements have evolved over millennia, through the mechanisms of identifying and eliminating or controlling the causes of illness /injury, as well as by identifying cures or disease control activities when illness and injury cannot easily be prevented.

Absolutely proving that something (an effect) occurs because of another thing (a cause) is not possible. Even if an apparent cause e.g. “bad air” appears to cause tuberculosis we cannot absolutely predict that it will always cause the disease. Even if in many, many instances it causes disease, this still does not mean that it will always cause the disease, as we cannot unequivocally say that in one or a few particular cases it might not result in disease. What we can do however is attempt to falsify the claim that it does cause the disease. If we cannot falsify that claim then we can reasonably assume for the time being that it is a true claim. If we can falsify it e.g. people who do not come in contact with the “bad air” also get the disease, then we reject the claim. We then have to look for an alternative explanation (e.g. bacteria in the air cause the disease) and attempt to falsify that claim. Again, if we can’t falsify it then we assume it to be true. If we can falsify it, then we look for an alternative explanation. There is therefore no absolute proof of causality or absolute truth. We can however get closer and closer to the elusive “truth” with each increase in knowledge over time.

Disease or ill health are usually caused by several factors rather than by one specific factor. So for instance people exposed to the TB bacteria (mycobacterium tuberculosum), do not get ill unless they have low immunological resistance to the bacteria, or increased susceptibility to it. Low immunological resistance could be due to poor nutrition; decreased immunity due to drug treatment; decreased immunity due to other infections such as HIV; decreased immunity due to cancer; etc. Increased susceptibility could be due to increased exposure due to over-crowding, or increased susceptibility due to silicosis.

Hence one needs a combination of factors (or causes) to be present before disease occurs.

However one of the factors is absolutely necessary and without it being present the disease does not occur. In the example above that **necessary factor** (or **necessary cause**) was the TB bacteria. This necessary cause is however not sufficient to cause disease, as it requires other factors (causes) to be present to result in disease. Therefore to develop a disease one requires **sufficient causes** to be present.

Occasionally only one factor (or cause) is required to cause disease e.g. rabies virus and haemorraghic fever virus. This means that it causes disease amongst all people who contract it. This necessary cause is then a sufficient cause, by itself, to result in disease. It is rare for causes to be both necessary and sufficient.

The requirement for sufficient cause to usually be composed of various component causes, allows one to group causes in different ways. A common way of grouping component causes (especially when considering infectious diseases) is to group them into those related to the Disease Agent (the necessary cause e.g. cholera bacteria), the Host (component cause could be e.g. malnutrition) and the Environment (component cause could be e.g. inadequate sanitation facilities). It is likely that combinations of various component causes in these groups would constitute sufficient cause. Several types of combinations amongst these 3 groups could constitute sufficient cause.

To complicate matters even further sufficient cause can change based on the availability of interventions. So, for instance in the example above, rabies virus used to be both a necessary and by itself, sufficient cause for rabies disease, since everyone infected with rabies virus got the disease. Then however an effective vaccine against rabies was developed, so sufficient cause for rabies is now composed of 2 component causes, namely exposure to rabies virus and not having being vaccinated against rabies virus.

Since there are many combinations of causes which can constitute a sufficient cause for illness or injury to occur, there exists a variety of possibilities for grouping or classifying causes of ill health. But why should one bother to group causes? Why not simply leave them as being various combinations of component causes which give rise to sufficient cause? The answers are that classifying causes assists one to advance a particular way of understanding why the illness arose and it is useful as a means towards taking action to prevent or manage a particular set of sufficient causes. A classification system must include all possible causes of an illness therefore all classification systems have sub-components. Classification systems therefore are devised for a particular purpose/s and the sub-components within them aid understanding of that purpose as well as contribute towards fulfilling the purpose. The number and type of classifications systems and the sub-components of the classification systems are limited only by the degree of insight/imagination and are linked to implicit/explicit ideological perspectives as well as the purposes they are intended to be used for. To make this more practically understandable several classification systems are summarised below with the IUR system explained in more detail than the others.

2.1 Immediate Underlying and Root causes of Disease Classification System (IUR System)(Also called proximal distal causes of disease classification system)

This classification system is based on interlinked levels of causality. **Immediate, Underlying and Root causes** are the 3 sub-components in the classification system. Each sub-component represents a particular level of causality which is inextricably linked to the other levels of causality. This system partially follows the necessary and sufficient cause logic with each of the levels of the causes requiring to have sufficient cause to produce an effect. It however rejects the idea that one single cause can constitute both a necessary and sufficient cause for ill health, as it asserts that there is always something deeper (the underlying cause; also called the upstream cause; also called the distal cause) that caused the immediate cause (also called the apparent cause; also called the superficial cause; also called the proximal cause) and hence the immediate cause by itself can never be sufficient. Similarly there is always something deeper causing the underlying cause which is the root cause (also called the basic cause; also called the base cause; also called the far distal cause). However where the two classification systems intersect is that each of the layers in the IUR system (Immediate, Underlying and Root) does not have to be a single cause. Therefore the immediate cause could be a mix of causes and those causes would be caused by a mix of immediate causes with that mix of causes then constituting **sufficient immediate cause**. However, that sufficient immediate cause would then itself still be caused by and underlying cause or a **sufficient mix of underlying causes**. Similarly then the underlying cause would then be caused by a root cause or a **sufficient mix of root causes**. This means that although there can be a **necessary immediate cause** and even a necessary and on its own sufficient immediate cause BUT that is only for the immediate cause and not for the linked chain of causes as that necessary and on its own sufficient immediate cause would still have underlying and root causes. So for example to the extent that viral haemorraghic fever is caused by a group of viruses which on their own are able to infect anybody the virus infection is a **necessary and sufficient** **immediate cause** but the underlying cause which dictates who gets exposed to the virus and who does not is still operative. Hence although the virus causes viral haemorraghic fever, it is the people who are deprived of land to live and farm on who are driven deeper into the jungle with increased contact with bats and monkeys who get infected, it is the people in overcrowded urban areas who have difficulty in isolating infected people and who lack the knowledge about the need to avoid even minor touching of the infected person who get infected, and it is the health workers in poorly resources health facilities who lack sufficient protective equipment who get infected when treating the infected people.

This IUR system provides one with insight into how causes are linked to the prevailing political-socio-economic system and what needs to be done at various levels if one hopes to implement substantial improvements. A further value is that it goes beyond the more obvious causes to uncover those factors that might not appear to be causal factors, but are in fact inextricably linked to the causal chain by virtue of being the underlying and root factors from which the more obvious and visible causes flow.

The use of the Immediate, Underlying and Root (IUR) system to group causes is in order to make clear the link between social justice and disease/ill health. When simply grouping disease into some ordering of sufficient causes, the factors (causes) which give rise to the sufficient causes are often forgotten or ignored. When using the IUR system one incorporates the concept of necessary and sufficient causes but adds in the component of social justice or equity.

When using this classification it is implied that one would consider **every** **disease or ill health condition** to be caused by a combination of Immediate, Underlying and Root causes. Therefore when dealing with the disease or ill health condition, one should attend to the full range of causes if one hopes to effectively tackle it. The sub-components are explained in more detail below.

**Immediate Causes**

The immediate cause is the apparent cause of the disease. It would constitute the necessary cause without which the disease would not occur. In infectious disease it would be the infectious agent. In traumatic disease it would be the trauma sustained. In chronic disease it could be a dietary agent, chemical agent, infectious agent or any risk factor for the disease. The immediate cause links with other underlying component causes to form sufficient cause for the disease to occur.

**Underlying Causes**

The underlying causes are those factors that in combination with the immediate cause make it possible for the disease to occur, or the factors which make the disease more severe, or the factors which make it more likely to come into contact with a necessary cause, or the factors which make it less likely to get assistance against the disease. Most of these underlying causes are intentionally differentially distributed amongst different socio-political-economic-religious groups, due to differential levels of wealth and power. Some underlying causes arise independently of group power dynamics, due to geographic location and natural disasters, but even these causes, as well as the effects they give rise to, could be ameliorated or enhanced by group power dynamics.

Underlying causes make it more likely that people will come into contact with both immediate and various other underlying component causes of the sufficient cause. Underlying causes have severity response mediation, with the potential to turn what would in many instances be a minor illness into something much more severe. Included in underlying causes are all those factors that would prevent or inhibit people from accessing health services. This would include access to treatment and rehabilitation for the disease, as well as preventing access to and uptake of promotion and prevention activities for the disease. Some underlying causes may be partially due to a direct result of one or more other underlying causes. This means that there are over-lapping layers of underlying causes and effects. Underlying cause “A” may combine with component underlying cause “B” and immediate cause “C” to form sufficient cause to give rise to effect “D”, but underlying cause “A” may also in combination with other component causes give rise to underlying cause “B”. Although underlying causes and immediate cause constitute sufficient cause, eradicating the underlying causes is usually not possible unless the root cause or causes are addressed.

**Root Causes**

These are the factors that result in the underlying causes being present amongst particular groups, but not amongst other groups. Some groups are therefore exposed to underlying causes while others are not. Those exposed to more underlying factors would then clearly be more likely to develop disease, as they would more easily have sufficient combinations of component causes to result in a sufficient mix of immediate causes to cause the disease/ill health. This is clearly unjust and inequitable. This injustice and inequity are perpetuated by differing power, control and violence levels within society. The propensity for some groups to maintain privileges for themselves at the expense of others (either consciously or not), is the factor which maintains the root causes in place. If one wishes to eradicate the root causes, one would have to oppose the view that some should be privileged above others and substitute the concept of equity or social justice in its place. One would then have to put in place structures and systems to ensure equity. When doing or attempting to do this, one would necessarily be in opposition to those who oppose equity and wish to maintain their inequitable and unjustly dominant position.

The root causes, while being far removed from the immediate apparent cause of disease, are therefore the ultimate cause of the disease. Eradicating root causes can only be done by fundamentally challenging the socio-political-economic structure of society.

While root causes imply that each cause is a basic essential cause, actually some root causes may be partially due to a direct result of other root causes. This means that there are over-lapping layers of root causes. This also means that there must be a fundamental cause to which all root causes could be traced. This fundamental cause is postulated as being a combination of greed, injustice and inequity.

2.2 Other Public Health Classification Systems of Causes of Disease

The essence of a classification system of causality is that all causes can be classified according to it (i.e. no causes are outside of the classification system). This means that the number and type of classifications systems and the sub-components of the classification systems are limited only by the degree of insight/imagination, and are linked to implicit/explicit ideological perspectives as well as the purposes the systems are intended to be used for. In general the intended purpose of the classification system determines the character of the classification system. Listed below are several commonly used classification systems.

**Preventable Level**

A classification system based on whether causes are preventable or not. Sub-components would be easily preventable, preventable with difficulty or not preventable. Causes that are not preventable are then further divided into whether they can be attenuated (easily and with difficulty) or not. The value lies in identifying those causes which one could easily attend to (by preventing or attenuating) and thereby concentrate ones efforts on those, as they are likely to result in greater benefit. Conversely it also shows which causes one should not devote energy to as they are unlikely to be able to be removed or attenuated.

**Prevalence Level**

This is a classification system based on the degree of prevalence of the causes in a particular location (country, region, district). Sub-components would be high prevalence, low prevalence and rare. The value lies in knowing which causes are most important for the population concerned and therefore what priority should be accorded to attempting to remove or ameliorate them.

**Burden of Disease**

This classification system extends the Prevalence Level system by considering not only what the prevalence of causes of illness are but also what portion of the total burden of illness in a community they directly account for. Sub-components are high burden of disease and low burden of disease. Since it uses a quantitatively measurable method to assess the actual proportion of the total burden of disease (as a percentage), it can more precisely determine the burden of disease instead of grouping them broadly into high and low burden prevalences.

**Cost Benefit Level**

This is a classification system based on the cost effectiveness of preventing or managing particular causes of illness. The sub-components are, in order of desirability, low cost with high benefit; high cost with high benefit; low cost with low benefit; high cost with low benefit. It strongly and rationally promotes that low cost with high benefit interventions should receive the highest priority. It is in essence a pragmatic mixture of the Preventable Level, Prevalence Level and Burden of Disease systems with an added specific quantitative economic component. The drawbacks of this classification system is that it promotes efficiency but does not address equity or social justice at all and may, while improving efficiency, actually worsen inequity. This is so because it does not identify who in a community benefits, or who has to bear the cost. As with the Burden of Disease it uses a quantitatively measurable method to assess the cost benefit value, and therefore it can be grouped quantitatively (actual cost benefit figure) in addition to the four sub-components.

**Equity Lens**

This is a classification system based on whether causes are linked to social injustice or inequity. Causes are classified into those that are unfairly distributed (inequitable causes) and those that are not unfairly distributed (equitable causes) amongst different social groups. Social groups could be gender, race, class, geographic area, caste, religious, age, occupation, etc. The benefit would lie in being able to establish that inequity exists, measure the degree of inequity, suggest ways of reducing the inequities, advocate for inequity reduction and actually implementing inequity reducing strategies.

**Aetiological Action**

This is a classification system based on aetiological mechanism or the way in which the cause results in the disease. The sub-components of this would be the various manners in which the cause are identified or the manner in which they exert their action to result in a disease. Sub-components could be chemical, biological, congenital, socio-economic, health system related, behavioural, infective, iatrogenic, toxin induced, trauma induced, environmental or idiopathic. This is in fact the classification system most commonly used by clinicians when they attempt to ascertain a cause of a disease in individual patients. The sub-components in this system are a mix and match and depend on the level of knowledge regarding the aetiological mechanism of action. The value lies in being able to design and implement appropriate interventions and being able to identify at which point one should target interventions.

**Necessary and Sufficient Cause**

This classification system is based on whether combinations of causes are sufficient to lead to an effect. Sub-components are necessary cause (presence is required) and sufficient cause (minimum combination of causes which by a combination of interactions would result in the effect occurring). The value lies in appreciating that causes are rarely isolated and are much more usually a combination of causes. Different combinations of causes would constitute sufficient cause to give rise to the effect. This allows for a more holistic view of causality and shows how different single causes appear to be more important in different contexts, because of the particular (different) combinations of sufficient causes that occur in different circumstances.

**Host, Environment, Disease Agent**

This is a classification system based on cause being a result of a particular interaction between the host, the environment in which the host lives and/or works, and the “disease causing agent”. This classification system arose from insights into why some people exposed to infectious micro-organisms developed disease while others did not. This concept has been extended to all ill-health conditions and not just infectious diseases. The value lies in understanding that a change in one or more of the above three aspects might result in an imbalance and consequently give rise to ill-health (e.g. individual host resistance is overwhelmed by the disease agent either because of low resistance [host factor] or strong disease causing effect [agent effect]). A healthy state would then be a state when the above three are in balance. This is in fact a particular way of applying the sufficient cause classification, where the potentially sufficient causes are put into these three sub-component groups. The virtue of this particular choice of sub-components is that it has the benefit of explicitly fostering a holistic understanding of how ill-health is caused.

**Level of Responsibility**

A classification system based on whether the cause is under the control of (or modifiable by) the individual affected or not. The sub-components would be modifiable, modifiable with difficulty and not modifiable by the individual. The value lies in being able to identify what causes individuals could prevent or modify using their own resources and initiatives. The logic here is that since it would have a beneficial effect for them they would be keen to act, and the only/main barrier preventing self help is lack of knowledge regarding these causes. This type of classification has fallen into disrepute as it often gave rise to a “blame the victim” effect and shifted the major portion of responsibility for low birth weight onto the individual, rather than it being seen as a group or societal problem.

**Public Sector Level of Responsibility**

This is a classification system based on which sector of government is primarily responsible for intervening to prevent or minimise the causes. It is similar to the system above but uses the government sectors instead of the individual. Sub-components would be causes that the health, education, social services, labour, housing, transport, etc., sectors are primarily responsible for attending to. The value lies in clearly identifying which sector of government is responsible for and should take the lead in attending to which causes.

**Task 1**

*Infectious diarrhoea is one of the commonest causes of morbidity and mortality in poor countries. Using the IUR classification system, list the causes of infectious diarrhoea under the headings of Immediate, Underlying and Root causes that you think would be present in your country.*

**Feedback on Task 1**

*Below (in short notes format) is a very comprehensive description of a wide range of possible Immediate, Underlying and Root causes that might pertain to various contexts in various countries. Your answer would not be expected to be as comprehensive as this, as you would have tailored your answer to the circumstances in your particular country. However some (perhaps most) of the various levels of causes below are to varying degrees likely to be applicable in your country.*

***Immediate Causes***

*Infectious diarrhoea is an infection that occurs when the gastro-intestinal tract is infected by viruses, or bacteria, or parasites that multiply in the intestine, directly damaging the intestinal lining and/or producing a toxin, both of which mechanisms result in the production of watery faeces (diarrhoea). The viruses, bacteria and parasites are present in faeces and sometimes in vomit. They are transmitted to the human hosts by contaminated water, by contaminated hands and by contaminated food. One or more of them constitutes the necessary immediate cause.*

(Note that diarrhoea may also indirectly occur in association with other diseases such as measles, upper respiratory tract infections and AIDS. Causes related to those infections that indirectly affect the gastro-intestinal tract will not be discussed further).

*Making up the mix of sufficient immediate causes would be one or more of the following:*

* *Lack of clean, infectious agent free drinking water*
* *Lack of adequate amounts of water for hand washing and basic hygiene*
* *Lack of adequate sanitation facilities*
* *Inadequate cooking of food*
* *Inadequate disinfection of contaminated water*
* *Lack of education about diarrhoea*
* *Malnutrition*

***Underlying Causes***

***(several of them are shown with their links to the immediate causes)***

1. *Lack of clean, infectious agent free drinking water is the main immediate cause of infectious diarrhoea. The infectious agents are transmitted in contaminated water. The underlying causes of lack of clean drinking water are:*

* *Lack of piped, chlorinated water due to lack of infrastructure*
* *Inability to pay for clean water, exacerbated by unemployment, part-time employment only, employment with low pay and high cost of water*
* *Contamination of wells, rivers and streams*
* *Lack of sanitation facilities (defecate near rivers and seepage through groundwater into wells)*
* *Heavy rain (washes infected faeces into rivers and increased seepage to wells)*
* *Lack of adequate storage of clean water*

1. *The underlying causes of lack of adequate amounts of water for hand washing and basic hygiene are listed below. (This lack of facilities to wash hands after defecation results in the increased spread of infectious agents via direct contact with other people and by contaminating food and/or water).*

* *Lack of piped, chlorinated water due to lack of infrastructure*
* *Inability to pay for clean water, exacerbated by unemployment, part-time employment only, employment with low pay and high cost of water*

1. *Lack of adequate sanitation facilities; no toilets or inadequately constructed toilets (people are then directly exposed to infected faeces). Lack of adequate sanitation facilities is due to:*

* *inability to pay for it,*
* *authorities not supplying it.*

1. *Inadequate cooking of food, so that infectious agents remain viable. This may be due to:*

* *Lack of electricity or unable to afford electricity*
* *Other fuels are more expensive or in short supply*
* *Time consuming to gather natural fuel sources*

*5. Inadequate disinfection of contaminated water because:*

* *Boiling water is expensive and time consuming*
* *Chlorination is relatively expensive*
* *Chlorine agents may not be available*

*6. Lack of education about diarrhoea. This results in:*

* *Poor hygiene practises even when hygiene facilities are available*
* *Inadequate and late treatment for the diarrhoea when it arises*
* *Lack of education on diarrhoea could be due to:*
* *General illiteracy*
* *General poor education*
* *General lack of health services*
* *Lack of health education provision*
* *Minimal emphasis by health departments on prevention and promotion*

*7. Malnutrition, which makes it more likely that people will contract infectious*

*diarrhoea and the resultant disease becomes more severe.*

*8. Migration and wide movement of people, where disease becomes more*

*widespread. Many rivers and wells in particular could become infected.*

*9. Commercial production and sale of food creates greater potential for infecting*

*large numbers of people via infected food. Larger volumes of food cause storage*

*problems with resultant contamination of food. Small number of infected people*

*can spread disease via poor hygiene practises to many others.*

*10. Lack of, or inadequate, health services results in greater likelihood of*

*contracting the disease and more complications from the disease as treatment is*

*not available. Inadequate health facilities are due to:*

* *Inequity in provision of health services*
* *High cost of health services*
* *Over-burdened health services*
* *Inaccessible health services due to location and/or opening hours*
* *Inappropriate health services with little emphasis on prevention/ promotion*

*11. Environmental damage due to heavy rains, strong winds and earthquakes*

*causes a spectrum of likelihood of contracting disease ranging from outright*

*disaster with destruction of homes, fields and infrastructure, to small scale*

*contamination of water sources and food.*

*12. Environmental damage due to drought results in scarcity of water with a*

*range of consequences from outright large scale famine to various levels of*

*under-nutrition. It also results in insufficient water for adequate hygiene*

*practises.*

*13. Epidemics of infectious diarrhoea in which rapid spread of some infectious*

*diarrhoea agents such as the cholera bacteria resulting in widespread*

*contamination of water sources and large numbers of infected people. This*

*facilitates further spread of the disease with a vicious cycle.*

*14. Inadequate or slow response by health services in preventing or controlling*

*epidemics. This is often due to:*

* *Poor infectious disease notification system*
* *Poor general health information systems*
* *Lack of planning for epidemic response*
* *Lack of concern about and lack of pressure on health services to respond*

*15. Cultural or religious practises which may predispose people to infection.*

*16. Congregation of large numbers of people (especially children) in one place at*

*crèches, schools or workplaces and gathering of people for various special*

*occasions often results in cooking large quantities of food which can easily*

*become contaminated.*

***Root Causes***

*There are many possible root causes of infectious diarrhoea.*

1. *High levels of Unemployment due to:*

* *Market driven economy (profit motivated and profit first) where people are treated mainly as labour units and surplus labour is desirable to reduce the price of labour.*
* *Employment skewed towards skilled and educated people*
* *Increased mechanisation*
* *Globalised economy*
* *Rapid (speculative) movements of finance*

1. *Unemployment which can result in:*

* *Poverty*
* *Dangerous and poorly paid work being accepted*
* *Malnutrition*
* *Increased exposure to disease*

1. *Structural poverty, which results in malnutrition and inadequate basic services with increased exposure to infectious diarrhoea and decreased ability to resist it. Structural poverty is due to:*

* *Global and National economic policy favouring economic growth over economic re-distribution and employment growth. This results in high levels of poverty for many and large amounts of wealth for a few.*
* *Race, class, religious and other types of group oppression.*

1. *Exacerbation of poverty by illness. For those who are already poor, illness is more likely to occur and may also prevent them from working, thus depriving them of income. Poverty is also worsened if illness results in disability.*
2. *Lack of access to land, which results in lack of ability to grow food or raise livestock, even if people have the time do so due to unemployment. This also results in lack of stability, lack of a home and lack of basic services as a consequence of homelessness or illegal settlement, due to a lack of property ownership or tenure. Lack of land for some groups is usually due to an excess of land accumulation by other groups, but in some areas could be due to an absolute insufficiency of arable land.*
3. *Deprivation of ability to thrive or even survive in this world due to the lack of:*

* *Land*
* *Means of cultivating land or living off it if available*
* *Employment*
* *Education or skills*
* *Any other means of income or access to resources*

1. *Poor education, illiteracy or low literacy levels, poor general knowledge of infectious and other diseases, and poor knowledge of infectious diarrhoea.*
2. *Racism. Deliberate oppression of particular racial groups results in higher poverty, lower education, lower self esteem, higher unemployment, lower paid employment, inadequate access to basic resources and decreased access to other resources. All of these factors contribute to higher levels of infectious diarrhoea as well as to many other diseases and ill health consequences.*
3. *Class or other group exploitation. Deliberate oppression as with racism; only the groups are different.*
4. *Inequity of health service provision. This results in fewer health promotion activities as well as less prevention and less treatment when ill.*
5. *Inequity of basic service (water, sanitation, refuse removal) provision. This results in increased risk of infection.*
6. *Lack of or inadequate social welfare system to assist those temporarily (or structurally) excluded from the economy and left with no means to support themselves. Clearly if there is structural exclusion from the economy, as exists in South Africa, then changes to the economic system are required, but an effective social welfare system could at worst assist in the interim, and at best be an integral part of those structural changes.*
7. *A lack of national policies to insist on equitable provision of basic services, or a lack of implementation of those policies.*
8. *Global economic system that favours wealthy industrialised and technologically advanced countries and groups of people, over poor countries and groups of people, putting profits before the well-being of people and communities.*
9. *Globally and/or nationally perpetrated severe, deliberate and open oppression of groups on the basis that they are lesser beings than other groups, e.g. slavery, racism.*
10. *Outright war or civil strife, with one group attempting to dominate another, but facing resistance.*
11. *Presence and perpetuation of greed, injustice and inequity in the world. Failure to realise that prosperity, well being and safety for one, rests ultimately on prosperity, well being and safety for all.*

2.3 Further exploration of Necessary and Sufficient component Causes of Disease

Now you may think that we have exhausted the subject of causality, but in fact we have just scratched the surface. The concept of causality is so important to the science and philosophy of epidemiology that it would be useful to pursue a more in-depth understanding of its complexities. The reading by Rothman and Greenland (1998) discusses the complexities of causality, starting with an assessment of sufficient and component causes, and explains how these complexities impact on our interpretation of the results of epidemiological studies. Go through this reading and then attempt the task below.

**Reading**

Rothman, K.S. and Greenland, S. (1998). Causation and causal inference. In K. S. Rothman and S. Greenland (1998). *Modern Epidemiology*. Philadelphia. Lippincott, Williams and Wilkins. Pages 7-17.

**Task 2**

Now that you have gone through this selection of texts on epidemiology, relook at the definition of epidemiology offered by Last (page 6 above) and then:

1. Write a short (half to full page) essay on whether you think Last’s definition adequately explains what epidemiology is, or not? If so, why do you think so? If not, why not?

2. If you were asked to supply an updated definition of epidemiology, what would you like included in your definition.

**Feedback on Task 2**

*Email your short essay and your updated definition of epidemiology to us and we will provide you with individual feedback.*

*For those who peeked at the feedback before answering the task ………. got you this time!*