

Research Proposal

Research & Development Methods

Table of Contents

Title	2
Background	2
Field of Study	3
Aim of Study	4
Expected Outcomes	4
Research Methods	5
Development Methods	5
Risks	6
Analysing Data	6
Resources	8
Bibliography	8

Introduction

Title

The title for this project is going to be “Statistical Mapping (Epidemiology Studies) of Measles outbreak in Swansea and South Wales”. It follows the news of the outbreak of Measles that has been a big part of media coverage over recent months, and would be an area of interest to the public.

Background

The area of study for this project is going to be in the field of Epidemiology, which is the study of the distribution of diseases throughout a population. Given sufficient data, it can allow the prediction of growth of a potential outbreak, and give patterns to allow government agencies to give useful information to the public. This information could help the public prevent the spread of disease, and possible re-emergence of another outbreak in the future.

The disease of concern with this particular outbreak is Measles, which is highly contagious and although young children are affected primarily, there is also a chance that adults can catch the disease as well. The symptoms (taken from Public Health Wales’ website) include a rash, fever, cold, fatigue and conjunctivitis. This is the disease that the study will be mapping.

There are a number of sources across the country that gather and display this information to the public. One example is Public Health Wales (a part of the NHS), which is in full use at the moment due to the most recent outbreak of Measles in the Swansea area and surrounding parts of South Wales. Other branches of public health exist across the UK, and Public Health England is also covering the outbreak of Measles as it threatens to spread further afield into the rest of the UK. Both of these branches operate in a similar fashion, giving out essential advice and information to people who are unsure about what the disease is (or the symptoms), any specific treatments that can be done, and the locations of local GP’s as to where they could get vaccinated. The data provided by the Public Health Wales website also offers the number of cases being reported by each local authority in Wales, and goes into further detail by giving out the times that the cases were reported.

Public Health Wales gives a background on Measles. It explains how it had been decreased to its lowest levels in the 1990’s due to vaccinations first introduced in 1968. The number of cases began its rise in 2006 when the first death from Measles occurred (1). It goes on to say that the outbreak rose in 2009 before decreasing to 8 in 2010 however in 2012 116 cases were confirmed in Wales. A BBC article with maps that contain data on the outbreak across the UK shows that during the first three months of the year 2013, 413 reported cases of Measles had been made in Wales (2). The article also makes clear that the outbreak is in the Swansea and South Wales area.

Field of Study

GIS can help visualise and analyse a variety of data in a way that people can understand. It has a wide variety of applications in the world, and has become a very useful method of displaying data in a readable way.

Geographical Information Systems allow the large quantity of information gathered in a subject area to be displayed in a way that is clear to the reader. There are a number of examples where GIS has been used, with one example to map the melting of ice sheets on each the North and South poles and in many other aspects of environmental studies. It therefore can be used to cover vast areas of land and data very accurately, with another example of environmental monitoring being the Amazon rainforest depletion rate. Weather is another key area for GIS, as data from satellites as well as factors such as temperature, and wind conditions are all taken into account to give us weather reports and forecasts daily. GIS can also be used to create the GPS systems used in many of today's cars, as well as giving authorities the tools to track a device if need be.

It has many different uses and epidemiology is another one of them, taking the number of cases from a particular outbreak and isolating the locations that the cases have been found. In fact the earliest example of GIS by John Snow shows the outbreak of cholera in the Soho area.

There are a variety of ways in which information can be displayed in geographical information systems, including graphs, maps and diagrams. Maps in particular will be useful for this project as they can display where the outbreak of Measles originated as well as the amount of cases to particular areas. Different timescales can also be used to compare and contrast the same areas at different points in time, and analyse this data to produce a projection of what may happen in the future. Other possible data that could be displayed is the difference between the numbers of cases in adults in children, which could display important information as children today may not yet have been vaccinated as opposed to some adults.

As mentioned GIS has been used in the field of epidemiology many times historically, and it will tie in very well with the title of this study due to the requirements for maps to be created that can show locational data of the outbreak, in a variety of ways. It requires these mapping techniques GIS can offer as the BBC article covering the outbreak has already demonstrated.

The study is topical as it is of interest to the public and could potentially affect a large percentage of the public. Due to this, the news media both in Wales and the UK as a whole are taking interest in what happens with the outbreak. As well as what happens as the outbreak continues, where it started is also of great interest to both the media and public as a result. This is crucial to potentially find out the cause of the outbreak and find a way to ensure that it does not happen again.

It is interesting as well because of the variety in which data can be processed and displayed. The different types of maps can give different perspectives depending on what is being shown and how it is being shown. While the different techniques of displaying data, such as graphs, can offer the same outcome of being able to manipulate data to show a certain trend. The data provided by Public Health Wales regarding the number of cases reported by each local authority allows the mapping of cases by each local authority's location.

Aim of Study

The aim of this study is to take all of the data available on the outbreak of measles and use the information in a number of ways through an application. The application should show the overall growth of the disease in the Swansea and South Wales area.

The objectives to achieve this aim are to:

1. Collect a substantial amount of data from government sources on the number of cases in the area, as well as the ages of people those cases that have been reported.
2. Design and develop an application capable of processing all of the data, which can analyse it and display it on a series of maps in a variety of ways. The programming language used to do this will be C#, and it will be a Windows Form application.
3. Make projections on how far the outbreak could spread, and how many cases there could potentially be in the future.

Expected Outcomes

The first outcome of the project should be a complete application capable of allowing the user to take data on the measles outbreak and enter it into the application. It will have a user interface that allows the different features of the application to be used easily. The features of this application will be:

- A login system to ensure that the data being used is protected.
- A navigation menu to access the various features.
- A data entry form to handle the input of the measles data.
- A page with options for displaying the data.

The information should be gathered from the reliable sources regarding the amount of measles cases that have been reported in Swansea and other areas of South Wales. There should also be enough data to be able to give accurate results after processing. The results should be able to give clear indications of how the disease has been spreading, and where it has been spreading to.

The application itself should be able to take the data it is given and apply it to a GIS information displaying method. The maps or graphs created by the application should display the data correctly and reveal a number of trends from the data it is given.

The application should also be able to give a projection based on calculations made while processing the data. These projections should be able to give realistic models of how the outbreak will spread.

Research & Development Methods

Research Methods

The project's time-scale will be longitudinal as it will measure the outbreak of measles over a period of time, extending throughout 2013. The data collection methods used will be a 'multi-method' of both sampling and secondary data. The secondary data will comprise of what is already available from government sources, and from this data samples will be taken. Gathering original data would conflict with the data that already exists and would be a time consuming and inefficient process as the number of cases of measles is relatively low in comparison with the number of people living in the Swansea and South Wales area.

The research strategy is going to be Action Research as the project will be able to add to the body of research already available through the use of GIS. This research will also identify trends that can be used as information regarding the spread of the disease.

This project will follow a research philosophy of positivism, due to the fact that it will involve a lot of data being collected. It will also involve obtaining quantitative results when the data is processed and displayed on any maps or diagrams. The results will also not be subjective as the data has already been gathered by sources such as Public Health Wales, as well as being specific and precise. The GIS component of the project strengthens its positivism as it is a very scientific approach. As mentioned however, there is some element of phenomenography due to the way different types of diagram can be interpreted when the information is displayed. Added to this are the projections, which will be estimations of how the disease will spread, and there are some aspects of this project which could be described as phenomenographical.

Due to the fact that the project is classed as positivistic, the logical research approach for the project is a deductive process. The same reasons that have been mentioned in relation to positivism can be repeated to describe why the project will be deductive. This means that the project will be focused around data and look to display trends in that data in a scientific way.

Development Methods

The representations of data such as graphs and maps are key parts of the project and what the application will be designed to achieve. Therefore the development of this application is going to need careful consideration and planning. Firstly, the application must be designed.

The design process will be to complete a series of requirements that the application must follow using diagrams such as use case and class diagrams, as well as use case documents and test documents. These documents and diagrams will describe each job that each part of the application is supposed to. Providing that the set of requirements is adequate, the application can then be created.

During creation of the application it will be important to keep running documentation on how the application is coming along, as well as what changes are being made to the design. After the application is created, the tests from the test document will be run to find out if all the areas of the

application are running smoothly. If the testing finds no errors, then the application can be completed. However if the testing picks up any bugs then the next step is to look at the documentation and try to rectify those errors.

Risks

There are a number of risks associated with any project, and it is no different for this particular project. The main risk to this project is the quality of the application that will be developed, as it needs to accurately process and portray the results of all the data collected. To avoid this, the programming of the calculations required must be accurate.

A much lower (but still evident) risk is the reliability of the data collected from sources. Even though the results of the outbreak can be found from government sources, we still may not be able to guarantee that each individual has given the correct information to the authorities. The entry of wrong data into the application will yield inaccurate results that will not be able to convey the correct location of the disease. As well as the small risk of the government data there is also the conflicting reports within certain outlets of the media that contradict the government data. This can be avoided by discarding the possibility of taking figures from media reports and focusing solely on the government sources.

Another potential risk is the loss of data from sources before all of it has been utilized. This could lead to incomplete and inaccurate results from the applications data processing, and to avoid this, a hard drive will store all of the data in back-up.

Another risk related to the loss of resources is the application itself, should the development be hindered in some way the entire project could be set back. This is where the back-up hard drive will be able to restore the application to a previously working state rather than needing to recreate whole sections of code at a time.

Analysing Data

The data must be represented in an accurate way; therefore the analysis must be completed to that effect. Once the data has been collected it must be sampled appropriately, this must be done by a number of categories. These categories are age and location, with location being the primary concern, as the sample must reflect that the majority of the cases in South Wales have arisen in the Swansea area. The proportion of cases should therefore be distributed in accordance with this fact.

Once the results have been collected and sampled appropriately the next step is whether the data is reliable. Data validation needs to take place to ensure that no incorrect data has made it into the sample. After the data has been checked there is the matter of what the data is going to display and what can be gathered from it. The results can be partitioned into categories similar to how the data was sampled to display the number of outbreaks that have occurred in children and adults. As mentioned the results will primarily show the location of the cases reported via maps with the GIS methods.

Simple calculations will allow the data to be brought down to small results such as the mean, which (when calculated) can show a single number regarding the likely age of the person. Mode can give

the most common location of cases once the data has been coded and is able to be calculated, but with the possibility of there being no mode should there be two locations that have the same number of cases. Depending on the trends shown in the results however, if the data is in some way skewed, the Median would be preferable to the Mean when it comes to finding out a mid-point in the ages of people with measles.

Other more complicated calculations can be made to find out the ranges of cases in certain areas of Swansea and South Wales, or a range of the ages of people that have reported having Measles. These would be the percentile ranges, which can reveal this sort of information. Another option would be the inter-quartile range that would achieve the same aim for a dataset that doesn't meet the requirements to use the percentile range.

These techniques will be used to find the trends in the data collected, these findings can be visualised in a number of ways. In chart form they can be displayed on a histogram, line chart or pie chart. These will give better understanding of the findings to the reader and show clearly the patterns related to this outbreak. Histograms will be used as a way of displaying the number of cases per certain age group, while pie charts are an effective way to display which part of South Wales has had the most cases during this outbreak.

Resources & Planning

Resources

There are a number of resources required to complete this project, firstly and most importantly, access to the data about the cases of Measles in the Swansea area. Without this there is nothing for the application to process.

The next resource would be a computer system that is well enough equipped to run Microsoft Visual Studio. Microsoft Visual Studio software is the next resource required to create the application that will help map the cases of Measles. Without an application, the data will not be able to be displayed in the maps that will show the trends of the spread of the disease. The language required to create the application will be found in Microsoft Visual Studio which is C#.

Another key resource is a back-up hard drive; to protect the data about the cases of Measles as well as keep a back-up copy of the application should the computer system lose either.

Bibliography

1. Public Health Wales | Measles. *Public Health Wales*. [Online] NHS, 11 April 2013. [Cited: 1 May 2013.] <http://www.wales.nhs.uk/sitesplus/888/page/43749>.
2. BBC News | Measles outbreak in maps and graphics. *bbc.co.uk*. [Online] BBC, 2 May 2013. [Cited: 2 May 2013.] <http://www.bbc.co.uk/news/health-22277186>.