Public Health GIS Unit
An Introduction to Geographical Information Systems for use in Health Care and Health Services Research

Table of Contents

1. Background – GIS in Health
2. Aims
3. Introduction
4. What is a GIS?
5. GIS Data
6. What Can a GIS Do?
7. GIS Use in Health Care Analysis
8. Implementing GIS
9. Problems with GIS
10. Summary
11. Directed Reading
12 References
## Unit 5: An Introduction to Geographical Information Systems for use in Health Care and Health Services Research

### Unit 5: Overview

An Introduction to Geographical Information Systems for use in Health Care and Health Services Research

1. Background – GIS in Health .................................................................3
2. Aims.................................................................................................3
3. Introduction......................................................................................4
4. What is a GIS?..................................................................................6
5. GIS Data..........................................................................................13
6. What Can a GIS Do?......................................................................17
7. GIS Use in Health Care Analysis.....................................................18
8. Implementing GIS...........................................................................20
9. Problems with GIS..........................................................................21
10. Summary.........................................................................................22
11. Directed Reading...........................................................................24
12. References......................................................................................25
1. **Background – GIS in Health**

The provision of health care in the UK has undergone radical changes in the last 10 years. However, while the delivery of health care has changed dramatically, societies concerns for the determinants of good health have changed very little in the last 150 years since Chadwick's pioneering 1848 Public Health Act. The issues of concern then, the inequalities in health, are the same as they are today (Calman, 1998).

Health inequalities are the disparity between health experience, and health care, that exist across socio-economic groups (from the most well off to the most deprived) and across geographical groups (the north and south) as well combinations of these within local communities. These issues have remained, though the focus of health care has changed. There has been a shift in medical emphasis towards primary and proactive health care services. The goal now is ‘preventative health care’ rather than the traditional ‘reactive health treatment’. Both the previous Conservative administration and the present Labour government have, through their respective White papers (Health of the Nation, and Our Healthier Nation) stipulated specific targets for health improvement in a selection of key areas. Both initiatives have placed considerable emphasis on the prominent role that information technology (of which Geographical Information Systems (GIS) are part) has to play in reaching these targets. Not the least of which will be the central role attached to GIS and spatial analysis in the functioning of the recently established public health observatories. One of the principle functions of the observatories will be health surveillance – a broad term that involves the tracking and monitoring of health. This complex and far from straightforward task will involve the assimilation and analysis of many diverse data ranging from poverty, housing, pollution, crime, educational standards; to pay, and employment, and not to mention directly related health data. These are all data relating to a geographical place, a common theme binding them together and one that explicitly places them in the domain of GIS.

The widespread use of affordable powerful desktop computers, and the availability of georeferenced socio-economic and health data has ensured that GIS are becoming widely recognised as powerful tools in health care planning, research and epidemiology.

2. **Aims**

The aim of this unit is to provide a basic but sufficient understanding of GIS, the specialist data used, and associated problems, to discuss the implementation of GIS and critically evaluate the use of GIS for health service planning and research.

2.1 **Outcomes/Objectives**

- Introduce GIS – what they are and how they can be used in health services planning and research
- Introduce the different data types used in GIS
• Identify some of the key problems with using GIS in a health care setting, specifically issues relating to data quality, data sharing and data confidentiality considerations
• Gain an understanding of the usefulness of GIS as an information tool
• Identify the key elements to formulate a GIS implementation strategy
• Direct you to further information

3. Introduction

We live in an information culture, and due to the rapid advances in information technology our access to information has never been quicker, or easier. The Internet gives us unparalleled access to information of such a diversity and magnitude that it would have been difficult if not impossible to comprehend even 10 years ago.

We have all this information, but where does information come from? This is a difficult question to answer both philosophically and logistically (i.e. what is information and where do we get it). However, irrespective of where we think the information has come from, information is nonetheless the product of other processes.

The raw ingredients of information are data: data are collected, collated, and finally disseminated. In other words data are searched out, found or arrived at; then put in order to make sense; before finally distributed for others to read and use without them having to go through the first two processes.

This unit is concerned with an analysis tool that is used in the process of turning data into information.

Data analysis is the process of turning raw data into meaningful information (see figure 1.). In the quantitative sphere it’s the combination of two elements:
• Statistical tools used to analytically sort and manipulate data, and
• Knowledge used to interpret derived results.

**Figure. 1 The process of turning raw data into information**

Statisticians that undertake statistical analysis need not be familiar with the subject area of the data they are working on. Their primary concern is the correct use of appropriate statistical methodology. Where as, the interpretation
of statistics on the other hand, and importantly much of the validity placed on results, depends squarely on the analysts experience and knowledge.

Often when results from studies are controversial, or hotly debated (pick any one of a number of recent headlines), the query generally is not from the statistical analysis - which is perceived as being grounded in sound mathematical principles, but instead on the subsequent interpretation of the statistics. The ability to place derived information into context is partially subjective depending upon the level of experience and knowledge of the researcher or analyst.

In an attempt to overcome the partially subjective nature of data interpretation, methods capitalising on high speed processing powers of computers have been developed. The aim is to derive as much information out a set of data as possible.

Some methods have focused on identifying patterns within the data. Often referred to as data mining or data trawling, the technology is based around advanced computing systems using artificial intelligence and neural networks.

An alternative approach is to place data into context to assist in its interpretation.

What does this mean? and why is context important?
Think pay rises for a moment!
You are suddenly told that your expected pay rise of 2% has dramatically increased to 15%. Wow. You think great, and you instantly start daydreaming with visions of that extra special holiday a planned assault on the shops, or a memorable ‘the drinks are on me’ night in your local.

You think you are, in short, 15% better off than you were.

However, this is the situation without context. When you get home and watch the news you discover that political instability in oil producing countries has dramatically cut the supply of crude oil. One of the most notable first effects of this will be the dramatic rise in the price of fuel. An increase in the price of fuel will directly affect not only the cost of the petrol you buy, but also the cost of products and services, the delivery of food, public transport, holidays, in fact it will affect the price of nearly everything. With this news the government announces that the estimated rate of inflation will now be 25% over the next year, instead of the planned 2%. Therefore, even though you have been given a 15% pay rise, the cost of living will increase by 25%, and so you are in fact 10% worse off. This is the situation after your pay rise has been put into context.

And so it is with health. Context is important for both the determinants of health, and the provision of health care. Therefore when planning the provision of

---

1 This is far from a tall story and occurred as a result of the military conflict in the Middle East in 1974, and that together with domestic political instability heralded an era of high inflation lasting until the end of the decade.
information for health management

Introduction for GIS for Health

Human activity is explicitly environmentally related; we do not live in a vacuum but continuously interact with our environment. Therefore the context in this instance is the prevailing environment around us, both in terms of physical attributes such as the land, air and sea (and the pollutants they may contain), and the non-physical such as our socio-economic and health status (our pay, diet, behaviour, education, and occupation etc.)

Environmental links with health were widely believed (if not understood) by the Victorians. They built a huge industry out of the belief that our natural environment could make us well. The growth of Spa towns such as Bath, Buxton, Cheltenham and Matlock allowed the rich to relax in comfort and ‘take the waters’ while the development of sea-side resorts such Blackpool, Skegness and Great Yarmouth allowed the less well off (if not the poor), the opportunity to escape the smog of the cities and breathe fresh sea air.

Self Assessed Question No. 1.
List as many of the recent reports as you can that have appeared in the national media that have purported to relate illness to environmental factors. Now do the same listing those health conditions that were not environmentally linked. Your first list is likely to be far longer than the second!

While this is not a particularly valid survey method, and in fact is more to do with how the media runs a story, it does nevertheless illustrate the prominent link between health and our environment.

GIS are one method of explicitly placing data in a contextual framework. The framework being the external geographical, environmental, or socio-economic world around us.

4. What is a GIS?

4.1 Background and History of GIS
It is perhaps easier to understand what a GIS is and what it does by tracing its origins. GIS developed through two parallel, but essentially unrelated activities, both dating back to the 1960’s. The first resulted from the increasing importance placed on emerging environmental issues and governmental recognition of the need to start managing natural resources (Marble et al, 1990). In Canada the process of creating a land inventory resulted in the mammoth task of dealing with records covering huge tracts of land and forestry areas – areas covering many hundreds of thousands of square miles. These initiatives occurred at the same time as emerging developments in computing and related systems (computer graphics and database managements systems; Jones, 1997). In particular the recognition of the potential offered by computers for the efficient and speedy data processing of large datasets matched the voluminous data requirements generated by the land inventory.
The considerable hurdles facing the Canadian’s, revolving around the need to combine traditional textual records such as land registries and cartographic land records such as maps with new database technology, led them to develop the first GIS – termed the Canada Geographic Information System (CGIS), (Tomlinson, 1990). Other systems followed to varying degrees of success (Marble et al, 1990). However, the use of the GIS in social science and health applications is still really quite new developing only from the early 1980’s (Scholten et al 1995, Martin 1996)

The significant breakthrough for increased GIS use in social and health research was brought about by the combination of the development of low cost high powered desktop computers together with the availability of socio-economic and spatial/environmental data (Martin, 1996).

GIS as a unified computer system results from linking parallel developments in many separate spatial data processing disciplines: Geography, Cartography, Remote Sensing, Photogrammetry, Surveying, Geodesy, Statistics, Computer Science, Software Engineering and Mathematics (Scholten et al, 1995). As a result of this diverse development GIS has been known by many names, depending on those who have used the system, and what it was used for, including:

- Automated Mapping and Facilities management (AM FM)
- Computer Assisted Mapping (CAM)
- Computer Aided Cartography (CAC)
- Computer Aided Drafting (CAD)
- Land Information Systems (LIS)
- Multipurpose Cadastre

The practical result of such a diverse user and development base is that there is no universally accepted definition of what a GIS is. There is however a more accepted view of what a GIS should do (Marble 1990, Martin 1996, Chrisman 1997). One common definition proposed by the Department of the Environment in 1987\(^2\), but since adopted by many encapsulates the principle features of a GIS:

“A System for capturing, storing, checking, manipulating, analysing and displaying data which are spatially referenced to the earth”

(Worboys, 1995)

\(^2\) Arising out of the Government report on the use and handling of geographic information in Britain, commonly referred to as the Chorley report (DoE 1987)
Essentially this means taking something that you can define with a physical location, and through its spatial coordinates, store it in the GIS, and then ask questions about it.

So what constitutes a physical location? This can mean anything from a house, road, river, telegraph pole, or an area of grassland for instance. These are ‘real’ things – they actually exist in a distinct observable geographic sense, but importantly within a GIS you can also incorporate ‘non-real’ things such as an area’s level of deprivation or its rate of crime for example. Such things have a very real meaning to us in an abstract sense, we understand what, and importantly, where they are, but they don’t exist as geographic realities.

It is the ability of a GIS to hold data about these less observable socio-economic phenomena, and in so doing make it useable for analysis, that makes a GIS so suitable for use in health care planning and research. For example, some aspects of the socio-economic environment such as the location of cities, the routing of bus services or the concentration of financial services in the City of London are directly observable. However other related elements of our environment such as poverty or ill health are not so directly apparent, but are nonetheless, very real and in existence, and what’s more they exhibit a distinct spatial distribution. That is, differences in the levels of poverty and health experienced by the population varies from place to place. There are no lines marked out on the roads defining these areas, but nonetheless people know which area is which, they know when they have moved from deprived to more affluent areas, though no one has actually crossed a real demarcation line.

Self Assessed Question No. 2.
Consider the socio-economic characteristics of the town or area where you live. You should be able to identify the more affluent and more deprived areas. You know where they are, but how would you define where they are. In other words, on a street map of your town, where would you draw the lines that demarcated the deprived from the prosperous areas? Have you chosen to do this following geographic features such as roads for example, if so why? What defines a community – is it the people in it, or the geographic space they inhabit?

4.2 GIS much more than just a mapping System
However, a GIS is much more than simply a storage and display system for what are essentially electronic maps. The heart of a GIS, and its distinguishing feature over other automated mapping and presentation systems is the ability of a GIS to manipulate and analyse spatially distributed data (Worboys, 1995).

A GIS is generally based around the same principles as a traditional relational database. However a GIS is more than just a database, a GIS is both a mapping/graphical presentation package and a database (see figure 2). Thus the data storage, manipulation and querying techniques associated with conventional relational databases, such as the widely used dBASE, Paradox and Access, are combined with the inclusion of a visualisation-mapping tool. Similarly a GIS can import and export data between spreadsheets and statistical modelling applications such as SPSS.
Figure 2. The main components of GIS: Database, Display and Computer

- **Database**: Technology underpins GIS. The requirement to store, search, and manipulate textual and geographical data are the defining features of GIS.

- **Powerful PC or desktop workstation**: The processing requirements of GIS are considerable, and until recently could only be undertaken using workstations. However, increasingly desktop PC’s have the power required to run GIS.

- **Graphical Display**: Large very high resolution display monitors are essential for GIS. GIS is as much a visualisation tool as it is a data processing tool.

The ability to run queries on data means a GIS can be used essentially in two ways. Firstly existing data can be displayed within a spatial framework (i.e. simply placing your data on a map), and secondly, queries can be generated that explicitly include the spatial framework. In this way hitherto ‘unseen’ geographical and environmental elements can be incorporated.

What does this mean? Firstly, taking data and placing it on a map. Observing the spatial distribution of mapped data can be appreciated in figure 3. The map here is taken from a study looking at the distribution of emergency (999 call) ambulance response times (Rivers et al., 1997). The mean response times were calculated and aggregated to electoral wards, part of the administrative geography in the UK (common units of analysis). The different ward response times are shown on the map ranked in three classes, with the darker wards showing the poorest times, and the lighter wards the better times – a typical choropleth map. Also included on the map are the locations of the three ambulance stations in this local authority area.
Providing a simple descriptive picture of the spatial distribution of your data can be very useful, as shown here, more information is provided than just the calculation of the mean response times. We can see here that, as might be expected, some wards close to ambulance stations achieve better response times. However it is immediately apparent that the ward of Youlgreave has a relatively poor mean response time, even though it is adjacent to a ward containing an ambulance station. It would be far harder to determine this with more traditional forms of data output such as graphs, tables and charts etc.

Using GIS has thus highlighted a factor in addition to the position of an ambulance station that may be affecting response time performance. The question that we now raise is: "why does the ward of Youlgreave have a poor response time even though it is adjacent to a ward with an ambulance station?"

In this sort of situation, the power of GIS can really be appreciated. The only other ward that achieves the same relatively poor response time to Youlgreave is Hathersage to the north. Estimating distances from ambulance station to ward might help explain the problem. Using the GIS technique of buffering, (see figure 4) equal distance circles radiating out from the ambulance stations are created (until the buffers meet). Here the straight-line radius of each buffer is 5 ½ miles. It is apparent that the vast majority of the ward of Hathersage is over 5 miles away from an ambulance station. So perhaps it is not unreasonable to assume that this influences the time taken for an ambulance from the nearest station to reach locations in this ward. However, the buffer also clearly illustrates that the ward of Youlgreave is well within the 5 miles distance, and so accordingly ambulance response times
should be equal to those around it\(^3\). However GIS is able to incorporate different datasets that might shed some light on the problem here. By incorporating major road networks one can identify a possible reason for poor performance in the ward of Youlgreave. Apart from a small section in the north, the ward does in the main contain no significant roads, those that do exist are relatively minor. Therefore, the sparse road network in the Youlgreave area may be an important factor contributing to poor response times.

5.5 mile buffers created around ambulance stations

All wards, except one in the 5 mile buffers have good response times. GIS has shown that the other poor response time ward lies outside of the buffer, \textit{i.e.} it is over 5 miles away.

It is immediately clear that there is only a small minor road in the north of the ward. Undoubtedly this will influence ambulance response times.

Using the example above, the kinds of questions one might want to ask when incorporating the prevailing spatial environment in your analysis are:

“What is the mean travel distance from ambulance station to incident?”; or

“What percentage of 999 calls are over X miles away from an ambulance station?”.

This type of study, simply displaying data on a map, applies little spatial knowledge. To assist in the interpretation of the maps, spatial data is included to put the original ambulance response time data into context. This spatial data, and the knowledge of what it means, allows a more informed interpretation of

\(^3\) The example here ignores prepositioning and ambulance ‘zoning’ which the author acknowledges are operational measures used by Derbyshire Ambulance Service (DAS) to address these problems. The intention here is to demonstrate GIS techniques only, and should not be construed as a criticism of DAS.
the distribution of your data. A GIS does this through the explicit use of spatial relationships.

A conventional map holds all sorts of information, the road networks, towns, rivers, and mountains etc. The relationship between these features on a paper map is implicit i.e. one can see where the roads pass through towns or mountain valleys for instance. It is therefore a straightforward process of plotting a route from town A to town B, avoiding busy cities and staying on ‘main’ roads. However these implied relationships have to be explicitly expressed if they are to be held in a GIS.

---

**Self Assessed Question No 3.**

Consider how you would describe to a friend (who has never visited before) how to get to your house.

You might say something similar to this:

Get off the M1 at Junction 33, and take the first exist along the dual carriageway. At the roundabout take the 4th exist to the housing estate. Follow the road for 5 miles, and take the left turning by the school. Continue along past the pub until you come to a cul-de-sac. My house is on the left by the big tree.

The above description is very easy and straightforward to follow. These are the kinds of directions we give every day. However, there is a very big assumption here. Consider for a moment what this is?

Think about the different geographical objects that are included in the above description: a Motorway, roundabout, housing estate, pub, school, etc. In addition a distance – 5 miles – has been specified.

These directions are only easy to follow if you know what these geographical objects are. The proverbial Martian from outer space would not know what these things are, so you would have to specify them. You might say that a road is a thin line. A school is not a line, but has shape that occupies an area. However, the above description just doesn’t include a road – but different types of roads, a Motorway, a dual carriageway and a cul-de-sac.

As you can see building up a picture of these objects, and how they relate and interconnect with each other can very easily become quite complex. A GIS database is different from ordinary databases because it is able to specify different objects and create ‘topological’ relationships between them.

It is the ability to define geographical features, such as roads, hold topological relationships about these geographical features, i.e. what bounds the road, such as other roads, towns it passes through etc, as well as the attribute data of the road, such as the road name, the class of road, how many lanes it has etc. that uniquely characterise a GIS, and distinguish a GIS from conventional databases.
**Self Assessed Question No.4**

An earlier section made the point that up and till very recently a GIS could only be used on workstations, and not PC platforms. Why do you think this is the case? - think about the data descriptions above.

Given the topological data relationships described above, GIS analysis can be computationally intensive, and until recently desktop PC platforms did not have the processing power required.

---

5. **GIS Data**

GIS is built around two data types:
- Spatial data – spatial objects - essentially maps
- Attribute data – the data about the spatial objects

5.1 **Spatial data**

Spatial data are the backbone of a GIS. Spatial data comprise the underlying geographic infrastructure such as roads, buildings, and boundaries etc. but also, as mentioned above, other less observable spatially distributed phenomena such an area’s level of crime for example.

Spatial data can be essentially classified in one of two ways. The difference between the two models is far more than a random choice depending on your particular preferences. The difference in the two models is quite important as it influences:
- The way in which you collect your spatial data in the first place
- The way in which you store and integrate your data within a GIS
- The way in which you undertake analysis

The two sorts of data are called Raster and Vector models.

5.1.1 **Raster Data Models**

Raster models assume that space is a tessellation or grid of cells. In effect a mesh of regular shaped and regular spaced ‘pixels’ (picture elements) are overlaid on your spatial objects. Each pixel holds one data value. So for instance if you overlaid your grid onto two fields, one of hay and one of grass, the pixels that fell on top of the hay might have a value of 1, and the pixels that fell on top of the grass a value of 2. In this way each row of the grid would be made up of either ‘1’s, or ‘2’s (see figure 5.)

A third value representing height can be included allowing three-dimensional modelling to be undertaken (the pixels changing to ‘voxels’), thus a mountain range can be incorporated within the GIS for instance. It is this three dimensional modelling based on a fine mesh of cells that forms the basis of Computer-aided Tomography – or CAT scanners – as this is an ideal model for representing gradual and very localised variations in your data, or for creating cross sectional views of your data (Jones, 1997).
Raster Data Model

The diagram on the left has two objects, one dark grey, and one black. The raster data model on the right depicts this by overlaying the area with a grid, and assigning a value to each cell of the grid. This forms a matrix of values each representing a single colour.

Raster data is collected mainly from remote sensed imaging – aerial photographs, or satellite images for example, it also forms the basis of desktop scanners and digital camera’s.

The second data model used in a GIS are termed vector models

5.1.2 Vector Data Models

Vector models assume an entirely different approach to that of the grid or raster based model. Vector models assume that everything can be based around the three geographic primitives of points, lines and areas (polygons). Where a point is a single feature represented by a single pair of x y coordinates, a line is the joining of two separated point objects, and a polygon is a closed set of line objects (see figure 6.)

The way in which a geographic object is considered a point, line or polygon depends upon geographic scale. For instance from an aircraft looking down, a house will appear quite small, no more than a point, however at street level the same house will obviously be much larger, with discernable detail, such as depth and area for instance, so it is likely to be represented as a polygon. Similarly from the sky a road may look like a single line, however at street level a road will have more discernible width with perhaps separate traffic lanes and so may also be represented as a polygon as well.
The main difference between raster and vector is that vector data uses specific coordinates for each location, rather than the value of a grid cell as with the raster. Therefore the vector model is considered to be the more spatially accurate, though vector data requires more storage than raster, and vector data are not as efficient at undertaking overlays – one of the main features of GIS based spatial analysis.

5.1.3 Attribute Data
Attribute data are the data on which analysis is undertaken. Therefore this data can be collected, manipulated and analysis performed in the same fashion as would ordinarily be the case when not using a GIS, but with one very important provision. The attribute data must have a georeference, i.e. it should have a link to a place in the real world, or at least, a link to a map. This can simply range from a set of coordinates, or more commonly from another dataset that has geographic coordinates embedded within the data.

Within health the postcode of the address in which a person lives is commonly used. There are a number of postcode spatial datasets available offering different degrees of accuracy. However they all work by representing a postcode (or an individual property within the postcode) by a set of geographic coordinates. Therefore if one was wanting to map the locations of, for instance hospital patients, then using each individuals postcode, they can, thorough this dataset, be linked to a place on a map within a GIS (see figure 7).

**Figure 7. Postcode Georeferencing**
Postcodes are extremely useful in mapping health and socio-economic data. A postcode represents on average 14 to 15 domestic delivery points (i.e. houses that receive mail), and is therefore an extremely high resolution spatial object. However, there are problems with using postcodes. Firstly, while the majority of the population have postcode not every one does, groups such as travellers, or people who live on canal boats for instance do not. Furthermore the Royal Mail updates the postcode file every 6 months to keep pace with changing and developing communities. So it can therefore be costly and time consuming to ensure the postcode file you have is accurate and up to date.

Self Assessed Question No. 5
Create a list of all the principle sorts of data that are used in your work environment (customers, patients, suppliers, contractors etc). When completed go through your list and identify which data or datasets potentially have a geographic dimension. Now think about how many of those data could be georeferenced (i.e. used in a GIS) using a postcode.

The data are combined in a GIS, and displayed in Thematic layers – each layer portraying a particular data feature. It is the principle of layering data, together with topological relationships, that forms the basis of GIS spatial analysis. See figure 8.

When viewed from the top looking down the layers are stacked on top of each other allowing the different data to be combined. A number of layers each representing different themes can be incorporated. These data may not at first bear any resemblance to each other, however, in a GIS, they are related by their spatial proximity.

Any point of interest, such as a patient represented by their postcode, can be analysed in relation to the other geographic features. For instance, where someone lives in relation to an area’s local industry. For example a GP clinic may notice an increase in patients presenting with respiratory disorders such as Asthma. GIS could plot where the people live to see if they are close to any potential causative pollutants, such as factory chimneys etc.
6. What Can a GIS Do?

Typically a GIS can be used to answer the following types of questions:

**LOCATION**
What is at .................?

All information about a place *i.e.* a hospital, its size, types of services provided

**NETWORK**
How do I get from.................to................? 

Best route between two locations according to some pre-defined conditions, *i.e.* rout planning for ambulance services for quickest travel time avoiding suburban streets with speed bumps

**CONDITION**
Where is.....................?

The best place to undertake an operation according to a condition, *i.e.* site a new GP surgery according to age structure of local population and access to local roads or bus services

**TRENDS**
What has changed since.....................?

Differences over time, *i.e.* examination of epidemiological data such as flu outbreaks in the same location over succeeding years

**PATTERNS**
What spatial patterns exist.....................?

Instances of Asthma between urban and rural locations

**MODELLING**
What if.....................?

Will the uptake for breast screening increase if a mobile service is introduced in rural areas.

(ESRI 1990)

Being computer based a GIS has the advantage of being able to model data and test ideas quickly and efficiently, as Nicol (1991) points out:

“The advantage in using GIS is the ability to test ideas quickly. Overlays, tagging of various scales of socio-economic data, and the modelling of relationships, are exceedingly cumbersome when done manually”

### Self Assessed Question No.6

Re-visit your answer to question number 5. Using your data consider which of the above GIS functions might be of use in your organisation.
The first process in outlining a strategy for implementing GIS is to consider the data used in your organisation. Firstly, determine if there is a spatial dimension to your work and your data. To do this, think about how the basic GIS operations (described above) could be used to assist in the kinds of data analysis undertaken by you or your colleagues.

7. GIS Use in Health Care Analysis


The use of maps in epidemiological investigation is far from new (Gilbert, 1958), with examples from the seventeenth and eighteenth centuries (Lawson et al, 2001). The significant increase occurred in the nineteenth century with the advent of methods in modern epidemiology and the systematic recording and collection of mortality and morbidity data. One of the most important and best-known examples of using maps for epidemiological investigation is the work of John Snow and his now famous study on the distribution of Cholera around the Broad Street water pump in the Golden Square area of Soho, central London (see figure 9 – taken from Gilbert, 1958).

During the London Cholera epidemic of 1854 Snow mapped each incidence of the in excess of 500 people who died of the disease during one 10 day period. In addition to the cholera deaths, Snow also mapped the location of each water pump from which the local population drew their drinking water. On examining the resulting map Snow noticed that the majority of deaths were clustered around one water pump located in Broad Street. Snow had the handle of water pump removed and observed that the incidence of new cases declined rapidly (Gilbert, 1958). It was discovered that unlike the other water pumps in the area, water drawn from the Broad Street pump was contaminated by raw sewage.

Figure 9. John Snow’s dot map of the September 1854 Cholera epidemic in Soho, London.

- Deaths from Cholera
- Water pump

Snow used a map to plot the deaths from Cholera. He found the majority clustered around the Broad St. pump. He had the pump handle removed and the incidence of new cases dramatically declined.
What made Snow’s study so important was his analytical use of a map. Prior to the 1854 cholera epidemic Snow had already postulated his hypothesis that poor sanitation could influence the spread of disease (Curtis et al, 1996). Furthermore, his method of mapping disease was in its self also not new (Gilbert, 1958). However, what set this particular study apart was the inductive reasoning Snow showed with regard to the transmission of cholera through the mapped distribution of the disease (Cliff et al, 1996). His conclusive demonstration, by cartographic representation, that cholera is water-borne, was the more outstanding and significant as it was made before Pasteur’s work on bacterial infection was published (Thomas, 1990). Such is the importance of Snow’s work that it is universally referred to and is considered to herald the beginning of modern epidemiology (Terracini, 1996).

Commonly the visualisation of the spatial distribution of health events in relation to background environmental data can be useful for gaining new insights, or indeed for hypothesis generation. As with the John Snow example, maps can reveal patterns not previously identifiable with data portrayed in tables and graphs.

GIS are able to undertake these tasks quickly and easily. This is the so called ‘added value’ that GIS can bring to traditional epidemiological investigation, as Wilkinson et al (1998 p182 ) comments: “The main purpose of disease mapping is usually to gain insights into the spatial distribution of disease determinants, or to develop hypotheses that may be tested using other methods.” Gahegan (2000, p79) suggests there are three aims to spatially visualising data:

- Promotes the discovery of inherent structure and pattern
- Prompts the generation of research questions and knowledge construction
- Enables the study of particular facets and dimensions in the data

The two principle areas of health services research where GIS is making a significant contribution is in resource allocation and planning (Paliwoda 1996, Foley and Frost 1996), and, spatial epidemiology (Thomas1990, Thomas 1992, Elliott et al 1996 and Gatrell et al 1998).

Within health a GIS can integrate the following data elements together:

**Health Infrastructure**
- Location of hospitals, GP Practices, Health Centres & Pharmacies

**Demographic Data**
- Age, Sex, Social Class and Level of deprivation

**Geographic Boundaries**
- Wards, ED’s, Community Nursing Areas, PCG’s etc.

**Health Status**
- Mortality rates, Morbidity rates
Health Influences
- Environmental (urban/industrial & rural areas), Behavioural (what people eat/levels of exercise)

Such data when integrated within a GIS can be used to assess:
- Where health providers are,
- Where risk areas are,
- Where the population is,
- Where services need to be.

For example:
- Is there a relationship between the demographic characteristics of a given area and the prevalence of certain health events or illnesses?
- Are certain services in higher demand in some areas compared to others?
- How will the closure of a health facility affect the local population?
- Do the services provided for an area match the needs of the local population?
- Where are the highest areas of deprivation?
- Where is the best place to site a new facility with regard to the target population for that facility?

8. Implementing GIS
As we have seen GIS is more than just another IT tool. GIS requires particular types of data, and given the intensive nature of GIS data manipulation and query operations a powerful computer. In addition given the visual nature of the output GIS require high resolution monitors (large ones at that) and very good printers. There is then the cost of the software itself, and finally not forgetting the training required of the operator. While modern GIS use window’s based interfaces, they are nonetheless sophisticated applications and require guidance and training to maximise their full potential.

Consequently, though GIS are very useful and most organisations will benefit from using them, certainly within the health sector, a clear plan and strategy for GIS use should be drawn up and followed before, during and after implementation.

Self Assessed Question No. 7
Draw up a draft GIS implementation strategy for your organisation. Start with establishing a requirement for using GIS. Critically evaluate how GIS could be used and what benefits your organisation will gain from using GIS. Then list the elements required to implement the GIS strategy.
9. Problems with GIS
GIS are very useful systems but they are not without problems. This in part stems from the two sorts of data used, attribute and spatial. Consequently the problems that exist with ordinary attribute based analysis can exist alongside the potential errors associated with spatial data.

9.1 General data problems include:
- Data are at the heart of any analysis - *rubbish in equals rubbish out*
- Getting data - Recall bias, selection bias, confidentiality and ethical considerations
- Lag time of disease presenting or symptoms becoming apparent from when actually contracted *i.e.* problems of population migration from one place to another. Someone could have previously been exposed to a causal agent, but who subsequently moved to a new area at the same time as the symptoms presented. Consequently if the new place of residence was close to a suspected environmental pollutant (though not the cause in this case) then potentially an incorrect association could be recorded.

9.2 GIS problems in particular include:
- Spatial data is costly – need to ensure it is up to date and accurate. If we use out of date spatial data such as old postcodes, then we could incorrectly georeferenced individuals. This is because after a period of time the Royal Mail re-uses a postcode, however it could be located in a completely different area to where it was originally. This could result in incorrect and inappropriate interpretation of data.
- Attribute data must be geocoded
The principle problem with using GIS in health care analysis is the availability of spatially referenced health data. As Mason (1994) points out the Health Service is often quoted as being the largest repository of locked up geographic data in the country, with the term *‘data rich and information poor’* frequently used.

However this problem is slowly being addressed after the instigation of a number of initiatives to increase the availability of health related spatial data:

**KORNER 1980** - Post coding health records

**BLACK 1980** - Linking health and deprivation in a spatial context

**CUMBERLEGE 1986** - Community health care based on location

**ACHESON 1988** - Health Authorities producing health statistics in a spatial domain i.e. maps

(Mason, 1994)

- Confidentiality and Ethical considerations – putting people on maps!
- Complexity of analysis – cluster analysis – think CJD, or Meningitis, determining if only a few cases of an outbreak in one area is a clusters is extremely difficult.
- Temporal considerations - modelling space and time. We move around in different places and different times.
- Home/Work/Leisure Journey Travel matrix – we do not exist in one place but constantly move around. Therefore is it appropriate to use a postcode that
represents just one place where we exist as our georeference. We may live
many hundred of miles away from where we work.

**Self Assessed Question No. 8**

Consider and evaluate the potential problems with using GIS in terms of
mapping individual patient data?

The main points to consider here are issues of confidentiality and disclosure of
personal information – does the use of GIS pertain to data protection legislation
or indeed European Court of Human Rights legislation?

---

The use of GIS in health applications can result in a paradox. The very power of
GIS is because it can accurately place individuals in their local environment. This
can include individual houses and roads. However in doing so ethical issues arise
because the very act of mapping someone can then potentially lead to disclosure
of private and confidential information. Consider sexually transmitted infections
(STI's). A public health department may wish to map its patients in order to see if
there is any pattern, or to target an intervention measure such as a health
promotion campaign. If the number of cases in a particular area are few (as they
tend to be with STI's) then a map could potentially allow identification of the
individuals concerned. The challenge therefore is to balance the need for more
accurate information against the risk of disclosure. This is easily achieved if the
results are kept internal and private, however, much of the public health function
revolves around communicating health information to the public.

---

10. **Summary** – see figure 10.

- Analysing data is a process of Converting Data---------Into---------Information
- Converting data into information (calculating statistics) is largely skills based
- Interpreting information largely rests on a knowledge and experience
- Placing data in context can assist in both deriving and then interpreting
  information
- Such a context is the prevailing environment and socio-economic world
  around us
- A GIS allows data to be placed in context
- A GIS is an information system *optimised* for handling spatially referenced
  data
- A GIS is both a mapping package and a database, but with enhanced
  capabilities
- A GIS allows the modelling and interpretation of different datasets
- A GIS uses two types of data: spatial data and attribute data
- There are two types of spatial data: raster (a grid), and vector (discrete
  coordinates) – the choice of which to use depends on your analysis
- GIS and spatial analytical techniques are not without problems Cost of data,
  confidentiality of data, complexity of analysis
- However GIS and Spatial Analysis is *VERY* useful at unmasking hidden
detail and *adding value* to your data analysis
Figure 10. Summary – the main features of a GIS

**GIS: an information system that links health and our environment through mapping**

**Health**
- Health research and planning are partly spatial disciplines:
  - Location of services
  - Health inequalities
  - Ecological studies
  - Disease mapping
- The focus is human activity

**Environment**
- Medical geography examines the spatial distribution of health:
  - Location of services
  - Health inequalities
  - Ecological studies
  - Disease mapping
- The focus is human activity

**Mapping**
- Health and geography joined by mapping, a common medium of communication and analysis.

**GIS**
- Mapping
- Computer
- User
- Database
- Graphical Display

**GIS is an information system optimised for handling spatially referenced data**

**Health GIS allows data to be placed in an environmental and socio-economic context**
11. **Directed Reading**

The list below contains some key texts relating to GIS in general and the specific use of GIS for health. For further references see the reference list.

### 11.1 General GIS

A superb overview of GIS can be found in:

This is an absolutely key text on GIS. It is derived from the so called ‘Big Book of GIS’:

details of this can be found at:
[http://www.wiley.co.uk/wileychi/gis/resources.html](http://www.wiley.co.uk/wileychi/gis/resources.html)

Kingston University GIS home page offers a good introduction to GIS, and links to further sites:
[http://www.kingston.ac.uk/geog/gis/intro.htm](http://www.kingston.ac.uk/geog/gis/intro.htm)

### 11.2 Health GIS

For examples of practical GIS use in health applications visit the Public Health GIS Unit website:
[http://gis.sheffield.ac.uk/unit/](http://gis.sheffield.ac.uk/unit/)

National Centre for Health Statistics – GIS and Public Health
[http://www.cdc.gov/nchs/gis.htm](http://www.cdc.gov/nchs/gis.htm)

This is an American site that contains a wealth of information relating to GIS use in the broad sphere of public and environmental health.

ESRI are one of the principle vendors of GIS systems. Details of GIS use in health can be found at:

### References


Mason, K. (1994) The Application of GIS to the Mapping of Medical Data for a Local Health Authority. SUC Bulletin 28 (1)


