



WOMEN'S MENTAL HEALTH & EMOTIONAL WELL BEING

Geographic Information Systems by Derry Well Woman in the
development of a Cross border Strategy for Women's Health

**A Feasibility study to test the appropriateness of using
Geographical Information Systems as a component in the
Derry Well Womens 'Health Policy Project'.**

Dr Adrian Moore

Dr. Sally Cook

Ms Dawn Fryers

Dr. Kerry O'Halloran

Ms Deirdre Fullerton

University of Ulster, Coleraine.

Contents

	page
Abbreviations	4
1 Introduction	5
2 Methodology	5
3 Results	6
3.1. Cancer data for County Londonderry	7
3.2. Cancer data for County Donegal	8
3.3. Cardio-vascular data for County Londonderry	8
3.4. Cardio-vascular data for County Donegal	9
3.5. Geo-referencing	11
3.6. Relevance of GIS for analysing the incidence and spatial distribution of cardio-vascular disease, and breast and ovarian cancer	11
4. Limitations and Recommendations	12
4.1. Limitations	12
4.1.1. Confidentiality of the data	12
4.1.2. ICD incomparisons	13
4.1.3. Modifiable Areal Unit Problem (MAUP)	13
4.1.4. Accuracy of geo-referenced data in ROI	13
4.1.5. Attendance at hospitals outside counties Donegal and Londonderry	13
4.2. Recommendations	14
4.2.1. Key fields in dataset	14
4.2.2. Temporal accuracy	14
4.2.3. Spatial accuracy	14
4.2.4. Base population	15
5. Conclusion	15
Appendix One	16
Data Questionnaire	17
Bibliography	19

List of Figures

	page
Figure 1. Steps to be undertaken to acquire a suitable cardio-vascular database for county Donegal.	10

List of Tables

Table 1. Relevant fields from the NICR database	7
Table 2. Relevant fields from the NCR database	8
Table 3. Relevant fields from the WHSSB database	9
Table 4. Relevant fields from the HIPE database	9

Abbreviations

CPD	Central Postcode Directory
DED	District Electoral Division
ED	Enumeration District
ESRI	Economic and Social Research Institute
GIS	Geographical Information Systems
GP	General Practitioner
HIPE	Hospital In-Patient Enquiry
ICD	International Classification of Disease
MAUP	Modifiable Areal Unit Problem
NCR	National Cancer Registry
NI	Northern Ireland
NICR	Northern Ireland Cancer Registry
NWHB	North Western Health Board
PAS	Patient Administrative System
PCN	Patient Control Number
ROI	Republic of Ireland
TRO	Tumour Registration Offices
WHSSB	Western Health and Social Services Board

1 Introduction

Derry Well Women commenced a two-year project in October 2000 on a 'Women's Health Policy and Implementation Strategy Process'. The plan of work to be undertaken for the project was divided into three phases. During Phase One, it was proposed, in conjunction with the University of Ulster, to conduct a Feasibility Study to ascertain the appropriateness of using a Geographical Information System (GIS) as a component within this project. The results from this would not only feed into the project in the short term but also provide a map for evaluating its impact in the long term. The objectives and aims of this feasibility study are outlined below:

Aims

- To conduct an investigation of the feasibility and usefulness of using GIS for the proposed study.
- To specifically target the incidence of breast cancer in two locations on either side of the border - Counties Londonderry and Donegal.
- To consider the availability of data to extend the above targeting to include ovarian cancer and to undertake a similar comparative analysis in relation to cardio-vascular disease as it affects women living in the same region.

Objectives

- To establish whether appropriate and sufficient incidence data exists in counties Donegal and Londonderry.
- To test the availability and comparability of relevant data sets in both jurisdictions.
- To establish the relevance of GIS as a means of mapping and analysing such comparable cross-border data.

This report focuses on the methodology, results, limitations and recommendations for the above aims and objectives.

2 Methodology

The methodology for the report focused primarily on meetings with the North Western Health Board (NWHB), Sligo General Hospital and AnPost and an investigation of other secondary data sources. **The emphasis on the Republic for meetings was because the research team have considerable knowledge and experience of data availability in**

the North. Telephone conversations were undertaken with personnel from Letterkenny General Hospital, the Economic and Social Research Institute (ESRI) and Western Health and Social Services Board (WHSSB). Relevant publications by the Northern Ireland Cancer Registry (NICR) and the National Cancer Registry (NCR) were also reviewed. This process of information gathering focused on the availability, comparability and quality of data with regard to cardio-vascular, breast and ovarian cancer in the two counties (Donegal and Londonderry). The purpose of this was to gain further insight into the data that is currently and likely to be available in the near future, and to determine if this would be suitable for use in a GIS.

To ensure the meetings and telephone conversations regarding the data were standardised, a data questionnaire was designed (see Appendix One) which included questions on the availability, quality and content of the datasets. The information obtained would enable conclusions to be drawn on accessibility and comparability of the data across the two counties and to assess the appropriateness and feasibility of GIS for analysis of the datasets for cardio-vascular disease, breast and ovarian cancer.

The questions served a variety of purposes. The 'dataset format' would confirm if the datasets from Northern Ireland (NI) and the Republic of Ireland (ROI) were comparable, and could be amalgamated together. The 'brief description of the dataset' determined the appropriateness of the data for the study. 'Confidentiality of the data' addressed any problems that may arise regarding acquisition of the data. Questions within the section 'data quality' give an indication of the reliability of the data. The 'data content' questions indicated the appropriateness of the data for implementation within a GIS. The 'geographical unit' (street address, townland, ward) determined the scale at which the data could be mapped and displayed. For example records available at postcode or street address level can be aggregated up to a variety of spatial scales such as Enumeration District (ED) in Northern Ireland or District Electoral Division (DED) in the Republic of Ireland.

3. Results

Brief descriptions of each of the datasets and the ease with which the data can be geo-referenced and mapped is discussed below (paragraphs 3.1 to 3.5). The relevance of GIS to analysis of the incidence and spatial distribution of cardio-vascular disease,

breast and ovarian cancer is also discussed in paragraph 3.6.

3.1. Cancer Data for County Londonderry

Breast and ovarian cancer data for County Londonderry is available from NICR, which was established in 1994 (replacing an older incomplete registry) with the aim of providing accurate and timely information on cancers occurring in the population of Northern Ireland. Data is collected electronically using the hospital Patient Administration Systems (PAS), Pathology Laboratory Systems, Registrar General's Office (Death Certificates), Central Services Agency, Ad hoc Registries and Radiology Systems. Upon amalgamating this data it is then checked for errors using International and USA programmes and also for duplicated records. Additional information from incidences obtained from death certificates is sought from General Practitioners (GP's). Data is available upon request, although the NICR has a strict code regarding security, confidentiality and issue of data. Names and addresses are collected for validation purposes, follow-up and survival analysis. However, ethical approval is required before named data is released for research purposes. The Registry has formulated a set of guidelines on confidentiality and issue of data, which is available on request.

Fields that are of relevance to this study and the percentage of completed records within each of these fields is outlined below:

Field	% Complete
Sex	100
Patient date of birth	100
Patient address at diagnosis (including non-melanoma skin cancer)	75
Patient address at diagnosis (excluding non-melanoma skin cancer)	79
Patient master index postcode (i.e. home postcode) (including NMS)	85
Patient master index postcode (i.e. home postcode) (excluding NMS)	93
Primary Site Code (initially ICD-9 translated to ICD-10) (Breast and Ovarian cancer)	100
Secondary Site Code (initially ICD-9 translated to ICD-10) (Breast and Ovarian cancer)	Not Available

Table 1. Relevant fields from the NICR database

Source: www.qub.ac.uk/nicr/research.htm

3.2. Cancer Data for County Donegal

The NCR was established in 1994 and aims to collect cancer incidence data (including breast and ovarian) from fifteen Tumour Registration Offices (TRO's) which are supplied with information from hospitals and other health care facilities located throughout Ireland. Funded by the Department of Health, this information is recorded on laptop computers and sent electronically to the NCR in Cork. Prior to analysis, data is checked for errors and duplicated records.

Again due to the confidential nature of address fields, ethical approval is required before they can be acquired.

A total of forty different fields of information are collected, some of which are extracted from the patient's medical records, allowing a comprehensive, complete and high quality dataset. Of the fields of information collected the following are of relevance to this study:

Field	% Complete
Sex	Not Available
Patient date of birth	Not Available
Address (four fields)	Not Available
Primary Site Code ICD-9 (Breast, 174 and Ovarian, 183)	Not Available
Secondary Site Code ICD-9 (Breast, 174 and Ovarian, 183)	Not Available

Table 2. Relevant fields from the NCR database

Source: www.ncri.ie

3.3. Cardio-Vascular Data for County Londonderry

Data on Cardio-Vascular disease incidence for County Londonderry is available from the WHSSB. The information available derived from the Patient Administration System (PAS). PAS data is derived from the records taken when the patient is admitted to hospital; the admitting nurse is able to check if the patient has previously been admitted to the hospital, thus reducing duplicates. Further data quality checks are undertaken by the WHSSB. The system covers a wide range of variables including:

Field	% Complete
Sex	Not Available
Patient date of birth	Not Available
Postcode	Not Available
Primary Code ICD-10 (from April 1996 prior to that ICD-9)	Not Available
Secondary Code ICD-10 (from April 1996 prior to that ICD-9)	Not Available

Table 3. Relevant fields from the WHSSB database

Due to issues of confidentiality there may be problems in acquiring the data, although it may be possible if it is aggregated up to larger spatial units.

3.4. Cardio-Vascular Data for County Donegal

The incidence of Cardio-Vascular disease for the county of Donegal can be obtained from Letterkenny General Hospital. This information is available from the Hospital In-Patient Enquiry (HIPE) database, which is derived from each patient's Whole Chart. Whole Chart includes admission and discharge letters, nurse reports, laboratory results and, if applicable, pathology reports. A wide range of fields is available and for this study the following fields are required:

Field	% Complete
Sex	Not Available
Patient Date of Birth	Not Available
Primary Code ICD-9 390 to 459	Not Available
Secondary Code ICD-9 390 to 459	Not Available
PCN (see below)	Not Available
Clinician/ Specialist	Not Available

Table 4. Relevant fields from the HIPE database

The reason for identifying the field Clinician/ Specialist, is that their consent is required for the field PCN to be included. This field is a code which is unique to the hospital, and matches up to the code on the patient's admission records. The admission records contain information on the individual's address, which is vital if the data is to be mapped. Unfortunately the patient admission records are still in paper format and as a

result this work would need to be **undertaken manually**. The process for obtaining data for Donegal is outlined in Figure 1.

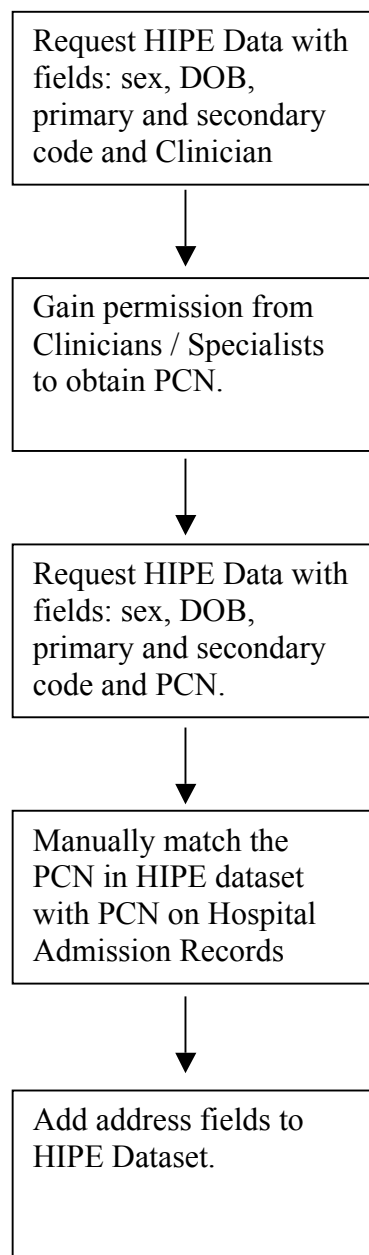


Figure 1. Steps to be undertaken to acquire a suitable cardio-vascular database for county Donegal.

Letterkenny General Hospital provide data for the Donegal region. If data is to be obtained on a National or Health Board basis, then ESRI would be required to amalgamate all the relevant data (this may have an additional cost implication which would be related to the extent of the work required). ESRI co-ordinate training and undertake further quality checks, however they would prefer the data to be sought from

individual hospitals if the PCN and address information is required. This could imply a considerable time delay in obtaining the data.

3.5. Georeferencing

To locate a case on a map, it is essential to obtain the grid reference (x and y coordinate) of the patient's address. Unfortunately the data from NICR, NCR, WHSSB and Letterkenny General Hospital do not include these fields. However, this information can be approximated from street addresses and postcodes. The Central Postcode Directory (CPD) is a United Kingdom based product which allocates an x and y grid reference to each postcode in a database. The system in ROI for undertaking this process is slightly different; Ordnance Survey Ireland and AnPost have created GeoDirectory, which assigns a grid reference to every address. This software is expensive to purchase and for the purpose of this study it may be more beneficial to contract this work out to a third party added value reseller such as Gamma in Dublin. The Geodirectory for Co. Donegal might cost in the region of £10,000 to purchase and would require software development to integrate with Board databases. To our knowledge the NWHB do not possess this data at present.

3.6 Relevance of GIS for analysing the incidence and spatial distribution of cardiovascular disease and, breast and ovarian cancer

A GIS, as stated by Burroughs (1998), is a powerful set of tools for collecting, storing, retrieving at will, transforming and displaying spatial data from the real world for a particular set of purposes. It is therefore hardly surprising that researchers undertaking studies related to the geography of health increasingly use GIS. This is also as a result of the increase in the accuracy of patient records, coupled with the realisation of GIS as an innovative and important component of many projects.

GIS is particularly well suited to studying epidemiological data such as the incidence data discussed in sections 3.1. to 3.4., as it offers display capabilities that assist the identification of areas with particular needs. Analysis of the spatial distribution of a particular disease enables the identification of areas with relatively high and low incidence rates. Mapping the incidence data also allows correlations and associations

with other spatial data to be made, for example demographic data. The user of a GIS can also analyse and report on variations in incidence rates over a period of time to assess the spread or containment of the specified disease.

A number of analytical studies of cancer have been undertaken in Ireland with the aid of GIS. A study undertaken at the University of Ulster examined the distribution of breast cancer incidence over the period 1993 –1995 in Northern Ireland, to investigate whether distance to primary and secondary level services is associated with differential incidence of the disease and stage at diagnosis. The National Cancer Institute (America) uses GIS Internet technology to illustrate both interactive and static cancer mortality maps, enabling breast cancer mortality rates to be illustrated over the internet (<http://dceg.cancer.gov/atlas/index.html>) by county and for two different time periods (1950-1969 and 1970 – 1994). A report on the incidence and mortality of cancer for all Ireland was published early this year, it contains maps illustrating age-standardised incidence rates for both breast and ovarian cancer by Health Board area.

4. Limitations and Recommendations

Limitations and recommendations which have been identified are outlined below.

4.1. Limitations

A number of limitations have been recognised which may have implications for acquiring and analysing the data.

4.1.1. Confidentiality of the data

Due to the confidential nature of the data and adherence to the Data Protection Act, problems may arise when trying to acquire it for research purposes. Ethical approval of the project will be required. This problem may be resolved if an agreement is made to ensure the patients' identities are not disclosed, for example by aggregating the data up to a higher spatial unit such as ED or DED. This would involve a researcher working in the Health Boards/Cancer registry and actually aggregating and anonymising the data before data can be obtained and any analysis undertaken.

4.1.2. ICD comparisons for North and South

Northern Ireland data is categorised using ICD-10 classifications, whereas ROI uses ICD 9. This should not pose a problem unless analysis of a specific sub-code of cardiovascular disease, breast or ovarian cancer is undertaken.

4.1.3. Modifiable Areal Unit Problem (MAUP)

A key component of all research is the spatial unit of inquiry. This can vary between the individual, the household, a village, a city, a Health Board or country. Much empirical analysis uses data aggregated to some form of geographical area. For example, the UK Census data is collected at individual household level and then aggregated up to a higher spatial unit such as ED, Ward or Local Authority. These spatial units, in say aggregated socio-economic circumstances may bear little resemblance to the specific circumstances of the people/households they contain. This is known as the Modifiable Areal Unit Problem (MAUP). Consequently, the analysis of data in different sets of zones or at different spatial scales, may alter the resulting pattern of aggregated observations. Openshaw (1984) states that associations between variables (for example morbidity and deprivation) are also prone to the impact of MAUP.

4.1.4. Accuracy of Geo-referenced Data in ROI

Representatives of AnPost came to Coleraine and provided a presentation on their GeoDirectory product. They did not however provide a live demonstration and as a result the accuracy of the results obtained by using Geodirectory cannot be commented upon. The claim is that they can accurately Georeference any residential household in the Republic to within a few metres of the centre of the property. Data is at the heart of a GIS and it is highly dependent on its accuracy. If the study is to be undertaken, it is recommended that the Geodirectory be used (either through direct purchase of the data for Donegal or by using a third party vendor) and that some form of quality assurance test be conducted on a sample of the records.

4.1.5. Attendance at hospitals outside Counties Donegal or Londonderry

Individuals who are residents of County Londonderry or Donegal and who have attended a hospital outside of these counties will not be included in the data. It was not possible to determine how significant this may be but is something that would need to

be addressed in a full study.

4.2. Recommendations

A number of recommendations which would facilitate analysis include the following:

4.2.1. Key Fields in Dataset

It is recommended that the following core items are included in each dataset:

- Sex
- Date of birth
- Postcode or address fields
- Primary ICD code
- Secondary ICD code

4.2.2. Temporal Accuracy

Temporal accuracy is particularly important if accurate conclusions are to be drawn from the data. It is recommended that the most recent datasets (for as long a period as possible) be obtained from the various sources to ensure that any analysis performed is accurate and reliable, and as a consequence facilitates effective decision-making. The data should also be collected for a similar range of years, for example, 1995 to 2001. It should be noted though that the most up-to-date address-based datasets may be a number of years old (e. g. up to 1997).

It is also extremely important to consider the total number of cases available for each medical condition and their geographical distribution as this will affect the ability to calculate robust, reliable incidence rates at small area level. Consequently, obtaining data for as many years as possible is of paramount importance.

4.2.3. Spatial Accuracy

As mentioned previously all data is available at street address/ postcode level, although for confidentiality reasons it may be difficult to acquire it at this level. While it would be much more preferable to obtain data in this form, because of the flexibility it provides in relation to spatial analysis, if the data must first be aggregated then it is recommended that it be acquired at the smallest spatial scale possible.

4.2.4. Base Population

To enable accurate results to be deduced it is necessary to standardise the data by taking into account the age and sex of the underlying population. This information can be obtained from the 1991 Census' for NI and the ROI (and 1996 for the ROI). Updates on this information, from the 2001 census will become available in the near future but it is unlikely that it would be in time for the duration of this project.

5. Conclusion

This report has aimed to ascertain the appropriateness of using GIS as a component within the 'Derry Well Women's' Health Policy Project with respect to investigating issues relating to the incidence of Breast and ovarian cancer and cardio-vascular disease. Through addressing the aims and objectives the following can be concluded.

Appropriate and sufficient data on the incidence of cardio-vascular disease, breast and ovarian cancer for counties Londonderry and Donegal exists, in a format suitable for use in a GIS, provided ethical approval is granted. Data from each of the two counties is comparable and could be amalgamated with a limited degree of editing. It would be feasible and extremely useful to use GIS for analysing the incidence and distribution of these diseases across the two counties. However, a number of matters relating to georeferencing, scale, spatial and temporal accuracy and data confidentiality would need to be addressed.

Appendix One

Derry Well Women Data Questionnaire

1. Availability and Purpose

Dataset title: _____

Dataset format: _____

Brief description of the dataset (field names): _____

Confidentiality/ availability of the data: _____

2. Data Quality

Source: _____

Geo-referencing accuracy: _____

Attribute data accuracy: _____

Temporal accuracy: _____

Data completeness: _____

Impact of missing data on analysis: _____

Data currency: _____

How easy is it to generate this as a clean dataset: _____

3 Data Content

Geographic units:

Street address

Postcode

- Enumeration District/ District Electoral Division
- Townland
- Ward
- District Council/ Local Government District

Aggregation unit: _____

Capture scale: _____

Coverage: _____

Number of records: _____

Bibliography

BURROUGHS. P. A. AND MCDONNELL. R. A. (1998) *Principles of Geographical Information Systems*. Oxford University Press, New York

OPENSHAW. S. (1984) *The Modifiable Areal Unit Problem*. Catmog 38, Geobooks, Norwich.

PROCTER. P. M. (1992) *Nurses, Computers and Information Technology*. Chapman and Hall, London

WALSH. P. M., COMBER. H., GAVIN. A. (2001) *All-Ireland cancer statistics 1994-96. A joint report on the incidence and mortality for the island of Ireland*. National Cancer Registry (Ireland) and Northern Ireland Cancer Registry.

Web Sites

National Cancer Registry (Ireland)

www.ncri.ie

Northern Ireland Cancer Registry

www.qub.ac.uk/nicr/intro.htm

National Cancer Institute

<http://dceg.cancer.gov/atlas/index.html>