

Actas del V Coloquio de Geografía Cuantitativa  
Universidad de Zaragoza  
1992, Zaragoza

**UTILISING POSTCODED ADDRESS-BASED  
DATASETS IN SOCIAL AREA ANALYSIS: A  
CASE STUDY IN NORTHERN IRELAND.**

Adrian J. MOORE  
Michael A. POOLE

*Department of Environmental Studies  
University of Ulster  
Coleraine  
Northern Ireland  
BT52 1SA*

**ABSTRACT**

The main purpose of this short paper is to provide a specific example of how address-based datasets can be utilised in social area analysis in Northern Ireland. The limitations of using conventional methods for the research project are outlined and the adaptation of a revised approach is presented and discussed.

**INTRODUCTION**

Just as the 1960s were considered by geographers as the era of the 'Quantitative Revolution', the latter years of the 1980's and the early 1990's are being accredited the accolade of being the era when "...geographic data handling had come of age."(Chorley,1991). A significant landmark for this development was the publication in 1987 of the report of a government committee into the 'Handling of Geographic Information', better known as the Chorley Report after the chairman of the committee, Lord Chorley(OPCS,1987).

The contribution of the report was to provide an overview of recent developments in information handling and to promote a greater awareness in public, corporate and academic circles of the value that can be made of an increase in the availability of new data sources. Allied to this was the pace of the development in computer based technology and the associated geometric decline in the costs of hardware and software products that could be used to handle this type of information. A primary technological development was the design of integrated systems and in particular database and graphical systems. Such systems are capable of handling, storing and manipulating geographically referenced data and are known under the collective umbrella of Geographic Information Systems(GIS).

One of the recommendations of the committee was that some form of standardised procedure should be adopted in the United Kingdom for referencing spatial data. They recommended that the Ordnance Survey National Grid and the Royal Mail postal coding system be adopted for this purpose.

## BACKGROUND

### The United Kingdom Postcode system

As in most other countries of Europe, the United Kingdom postal service(Royal Mail) devised a system of reducing costs and speeding up the handling and delivery of the mail. The postal coding system used is known as the 'postcode' which consists of a string of up to seven alphanumeric characters. Postcodes are hierarchical in nature and are composed of a combination of individual codes which identify addresses into discrete geographical units. The largest unit is the postal area which is sub-divided into districts, then sectors and finally into unit postcodes which contain on average 17 addresses or delivery points(Raper *et. al.* 1992).

In the United Kingdom as a whole there are approximately 1.7 million postcodes covering almost 24 million addresses and in Northern Ireland there are just over 40 thousand postcodes covering about 600 thousand addresses. Table 1 shows a breakdown of a typical postcode for a household in Northern Ireland 'BT60\_1BQ' along with the number of

each type of unit in the province(See also Figure 1 for the spatial distribution of postcode districts).

Table 1. Breakdown of a unit postcode and number of each type in Northern Ireland.

Postcode BT60 1BQ'	Type	Number in N. Ireland
BT	Area	1
60	District	81
1	Sector	232
BQ	Unit	41,203

Source : Royal Mail 1992 (Unpublished)

While the postcode has proved to be a very successful device for increasing the speed and efficiency of handling the mail it has also been adopted for many other useful purposes, most notably in marketing and research. By the very fact that the postal coding system is geographical and hierarchical, it can be used to provide a simple, easy to interpret, spatial identifier for people, households and businesses. A prime example of an application of the use of postcodes for a purpose other than the mail is that made by insurance companies in working out and reviewing car and household insurance premiums.

What facilitates this type of use is the availability of a number of postcode related products which are provided both by Royal Mail and some private companies. One of these products comes in the form of a computerised database of addresses and postcodes called the Postcode Address File(PAF). An additional product, which was developed from the PAF in conjunction with the Office of Population Censuses and Surveys is the Central Postcode Directory(CPD). Additional information such as local Authority ward code, National Health Service area code and grid references based on the British National or Irish Grid are provided. The CPD was produced for use by the Health Service and other government bodies but a machine readable commercial version of the file, called the Postzon file, has been made available by Royal Mail. It is the ability to combine the postcode with O.S. grid references and electoral wards which provides the greatest opportunity for

alternative uses of address-based data in the United Kingdom. The advantage of the Postzon file is that it provides the link between two incompatible forms of spatial representation, the postal system and the electoral system (Scotland is the exception, see Raper *et. al.* 1992. Also see Table 2).

Table 2. The availability of 1981 census material in the United Kingdom.

	England and Wales	Scotland	N. Ireland
District/Ward	Yes	Yes	Yes
Small Area Enumeration District	Yes	Yes (Postcode Based)	No
Small area Metric Based	No	No	Yes

For the social researcher, probably the main source of information is the decennial census which provides biographical, social and economic information about the national population. In Britain, data is provided publically in aggregated form using the electoral system of districts, wards and enumeration districts as the spatial units of measurement. This is not true for Northern Ireland though, where neither maps of enumeration district boundaries nor the related census data are available in any useable form (See Table 2). The ability therefore of the CPD to relate address based data to the census via the local authority ward codes makes it potentially a very powerful tool for handling geographic information in social research.

The system however is not altogether perfect. Some have expressed concern about the methods and accuracy of the postcoding, grid referencing and ward allocations in the Postzon file (Wilson and Elliot 1987, Gatrell 1989 and Raper *et. al.* 1992). The implications of these known levels of inaccuracy in research ultimately depends on the type and scope of each individual application. What follows is the presentation of a general methodology that was developed as a consequence of limitations in using the conventional methods of handling address-based data for a specific research project in Northern Ireland.

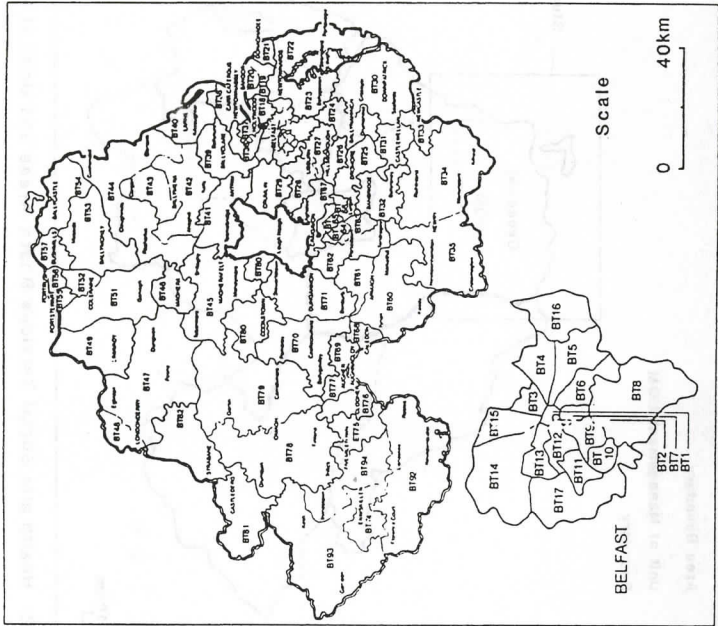


Figure 1: Map of Postcode Districts in Northern Ireland 1991

Source: N. Ireland Postcode Directory 1991

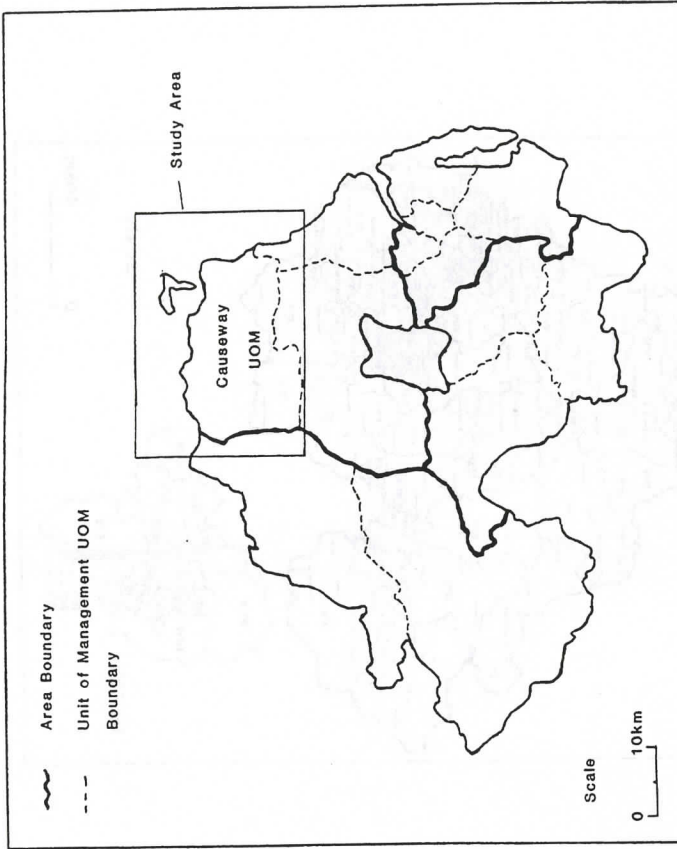


Figure 2: Health and Social Services Board Areas and Units of Management

## **The Research Project**

One consequence of the changes made in the organisation and provision of general practitioner services in the United Kingdom since April 1990 was a review of the payment system (Health Departments of Great Britain 1989). In addition to the basic capitation payment for number of patients on their lists, general practitioners can also qualify for a special payment based on the number of patients living in electoral wards which could be described as 'deprived'. Deprivation is measured according to an Underprivileged Areas Score (UPA) known as the Jarman Index (Jarman 1983). The index is a composite score of eight socio-economic variables derived from the census. The use of the word 'deprived' is somewhat of a misnomer as the measure was originally devised in the early 1980s as an indicator of general practitioner workload (Senior 1990).

A research project was undertaken by the authors to examine the effects of the socio-spatial characteristics of patient lists on general practitioner workload. The project had two general aims. The first was to describe the patterns of uptake of general practitioner services for a known threshold population, and the second was to relate these observed patterns to the Jarman index which purports to be a predictor of workload.

The primary data sources for the project were two address-based datasets which, for reasons of confidentiality, contained postcodes rather than names and addresses. The first was a directory of all patients (over 92,000) registered with doctors in the study area (Causeway Unit of Management, see Figure 2). The other was a sample of workload from general practitioners partaking in the survey (over 4,000 records).

## **METHODOLOGICAL AND TECHNICAL PROBLEMS**

### **Methodological**

The main problem from a methodological point of view was the unsuitability of using the electoral ward as an appropriate spatial unit for the measurement of deprivation. As Senior (1990) pointed out, from research he conducted in Great Britain, composite scores built up

of aggregated census data at ward level can be misleading. A primary concern is the heterogeneous nature of socio-economic variables within wards. The fear is that in using the ward as the unit of measurement, small pockets of deprivation would be lost both in the ward deprivation scores and in any subsequent analysis that would be undertaken. The problem may be less important in studies of large metropolitan areas where wards might be more homogeneous, at least for some socio-economic variables. The same cannot be said of smaller urban settlements where social area homogeneity is apparent at a finer scale level; a characteristic which is particularly true of Northern Ireland (Poole 1991). Rural wards pose an even greater problem for the researcher given the fundamental problems of trying to measure socio-economic status in rural areas.

Consequently, it was considered reasonable to examine the data at two levels, the first using the ward as the spatial unit of measurement and the second examining data from the urban areas separately. With regard to the latter, the problem arose of how to devise a finer and more meaningful spatial system of categorising the population of relatively small urban areas while at the same time being able to utilise the data from the census.

### **Technical**

From a technical standpoint the initial problem encountered was how to allocate the postcoded datasets to electoral wards and subsequently to Jarman scores (and census data). This would enable the researcher to build deprivation profiles for practices and analyse the workload data relative to deprivation. The easiest conventional method of performing this operation would be to use the CPD Postzon file for Northern Ireland. However, a number of problems were identified in trying to use this method.

The main limitation in using this dataset was the accuracy of the grid referencing of postcodes. With the exception of the Belfast metropolitan region, where postcodes are grid referenced at hundred metre resolution, all the other postcodes are referenced only to kilometre level. This would cause problems when trying to calculate accurate distances between patients and surgeries which are needed to examine the effects of distance on workload. It was anticipated that this would be particularly



important for urban areas where, in a previous study, one of the authors found marked differences with distance in the utilisation of primary care services in one of the towns in the study (Moore,1988).

There is also concern over the accuracy of postcode to ward allocations. Bond *et.al.* (1991) found a 52% error margin in allocating new 'imputed' postcodes to wards. The need for accurate ward allocations even for primary analysis was considered vital.

Finally, at the time of study the only version of the Postzon file available was from 1985. This version only gave ward allocations based on the boundaries which were devised in 1983/84 and therefore were incompatible with the 1981 census material. In addition, it was also noted that many postcodes in the survey data would not even be included on the file by virtue of the fact that they only came into existence after 1985. What was required then was a refinement, for at least urban data, of the type of information available from the Postzon file, namely grid references and 1981 ward allocations.

## **A REVISED APPROACH**

### **A 'Natural Areas' Classification**

The problem of creating a new classification of urban areas at a finer level than the electoral ward yet still making use of census material was overcome by the manner in which the census small area statistics are provided in Northern Ireland. As stated earlier, enumeration district data is not available in Northern Ireland in any useable form. However, data is available in a metric form at kilometre grid square level for the whole province and at hundred metre grid square level for most urban areas, even down to very small villages. The ability to access census material in urban areas at such a fine level provided a flexible framework from which new spatial units could be created.

Poole(1982) exploited this flexibility by creating an intra-urban regional classification for research into residential segregation in Northern Ireland. Using similar principles and guidelines, a set of 'natural areas' was built for the five towns in the study area, using amalgams of hundred metre grid squares.

In an attempt to maintain some form of internal homogeneity, the broad housing tenure characteristic of either public or private was used as a guideline. Tenure was chosen for two reasons, the first, because it is clearly segregated spatially and is easy to identify visually and from maps, and secondly because it has been found to be a good measure of socio-economic status related to health (Haynes 1990).

Two scale levels of regionalisation were adopted, a macro-level of 120-300 households (with an average of 200) and a meso-level of 60-150 households (with an average of 100). Meso-level regions nest perfectly into macro-level regions and each macro-level region contains one, two or three meso-level regions. An example of this system is given in Figure 3 for the town of Portstewart. It is immediately apparent that there is a distinct advantage over using wards as the spatial units of measurements in the study.

### **Grid Referencing**

Given the complexity of the problem and considering that the facilities of a GIS were available it was decided to try and refine, for the urban postcodes only, the kilometre square grid references that were provided by the Postzon file. With no maps of unit postcodes available, the local postcode directory was used to identify each postcode by town. From Ordnance Survey maps a more accurate hundred metre grid reference was obtained and a computer data file manually compiled. This way all new postcodes from 1985 to 1991 could also be identified and spatially referenced. Rural postcodes were omitted because it was considered that the benefits of obtaining, or trying to obtain, a hundred metre grid reference over a kilometre square grid reference for a rural postcode would be negligible given the low levels of population density in these areas. The kilometre square grid references provided by the Postzon file for rural postcodes were therefore accepted as the most accurate available. In this study it was not considered to be so crucial given that the hundred metre resolution would be of more importance in the urban part of the analysis.

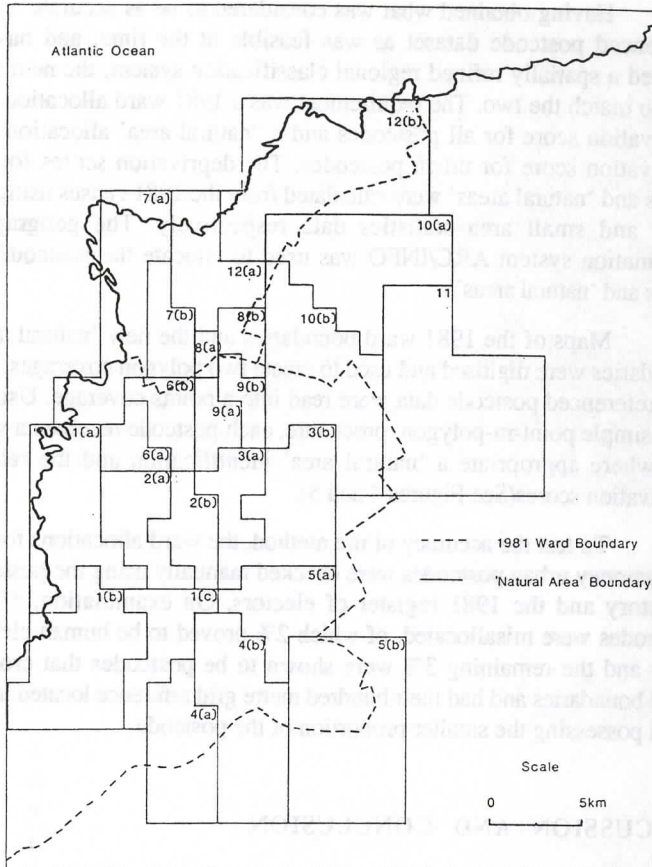


Figure 3: Electoral Ward and 'Natural Area' Boundaries for Portstewart, N. Ireland

## The GIS Contribution

Having obtained what was considered to be as accurate a grid referenced postcode dataset as was feasible at the time, and having devised a spatially refined regional classification system, the next step was to match the two. The requirement was a 1981 ward allocation and deprivation score for all postcodes and a 'natural area' allocation and deprivation score for urban postcodes. The deprivation scores for the wards and 'natural areas' were calculated from the 1981 census using the ward and small area statistics data respectively. The geographic information system ARC/INFO was used to allocate the postcodes to wards and 'natural areas'.

Maps of the 1981 ward boundaries and the new 'natural area' boundaries were digitised and used to create two polygon coverages. The grid referenced postcode data were read into a points coverage. Using a very simple point-in-polygon procedure, each postcode received a ward and where appropriate a 'natural area' identification and the related deprivation scores (See Figures 4 and 5).

To test the accuracy of the method, the ward allocations for the Ballymoney urban postcodes were checked manually using the postcode directory and the 1981 register of electors. On examination, 5% of postcodes were misallocated, of which 2% proved to be human clerical error and the remaining 3% were shown to be postcodes that crossed ward boundaries and had their hundred metre grid reference located in the ward possessing the smaller proportion of the postcode.

## DISCUSSION AND CONCLUSION

Considering the objective of the original study, it could be argued that the creation and application of a revised methodology has been a success, especially given the levels of accuracy attained. The basic approach of using conventional methods of utilising address-based datasets for the project were found to be unsatisfactory, thus presenting a difficult research problem.

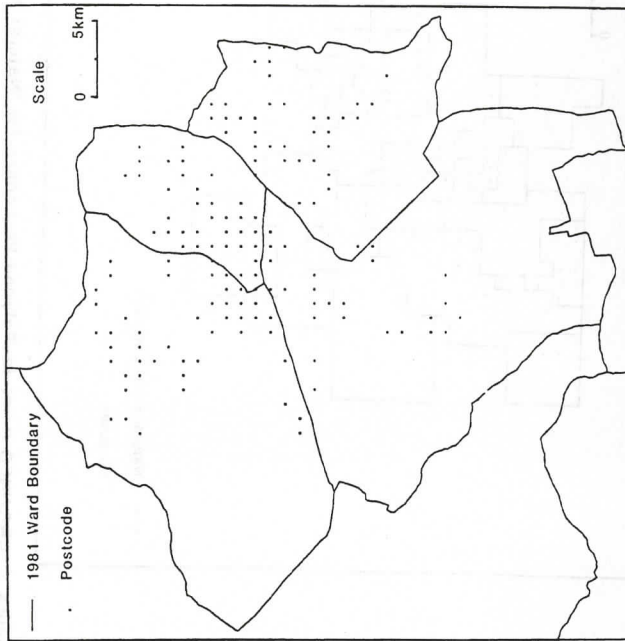


Figure 4: ARC/INFO Point-in-Polygon Procedure for 1981 Wards in Ballymoney

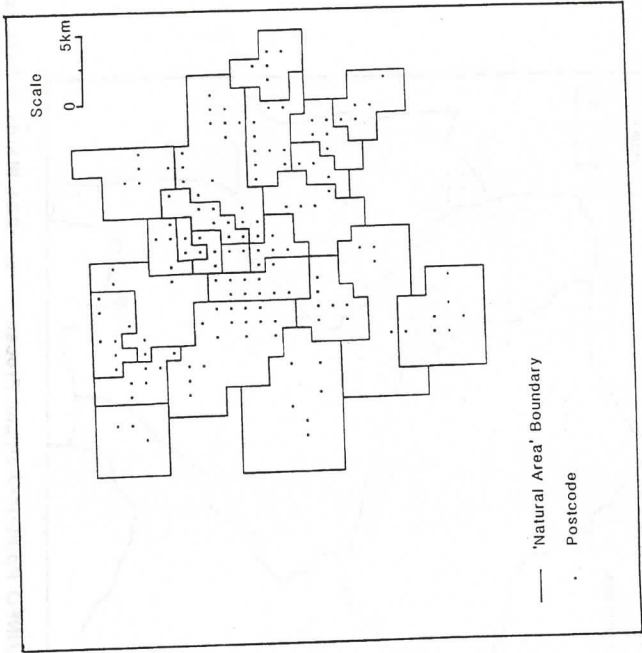


Figure 5: ARC/INFO Point-in-Polygon Procedure for 'Natural Areas' in Ballymoney

By investing some time and resources at a 'grass roots' level a revised methodology was developed. The key factor in the success of this methodology was the use of the geographic information system ARC/INFO. It provided the link between two, and subsequently three, hitherto incompatible forms of spatial representation, the postal coding system, the electoral system (wards) and the 'natural areas' classification system. The specific benefit of the linkage is the ability to use the vast amounts of social information available from the census. In addition, the power of the GIS, which was only touched on in this application, allows the researcher, in theory, to relate survey data to any other information that can be spatially referenced. This information can be either human based or physical, for example, information on morbidity and pollution data; the scope is vast.

The new methodology has other advantages as well. It can be used as a template for any social research application that requires the use of large address-based datasets and makes use of census material at ward or small area level. In addition, the use of a metric grid system at the small area level is beneficial in that it can make use of data from the most recent census in 1991 as well as previous censuses. The opportunity for longitudinal social study in urban areas of the province becomes more feasible.

The main purpose of this short paper was to provide a specific example of how address-based datasets can be utilised in social area analysis in Northern Ireland. It is also hoped that the paper can be used to illustrate a more general point concerning the quality of data input in this type of social research. The point is, that while the technology and computer based products may be available, it is the quality of the data sources being used which is crucial. If poor quality data are used then poor quality results are assured. It is essential, therefore, to consider the quality of available data in relation to the needs of the project, and where possible to update and clean that data. The consequence of better preparing input data is that available technology such as geographic information systems can be used to full effect in producing accurate and dependable results.

### Acknowledgements

The authors would like to thank Brendan O'Kane, Nigel McDowell, Kilian McDaid, John McKenna and Joanne McDowell for their help in preparing this paper. The financial help of the Department of Education for Northern Ireland is gratefully acknowledged.



## REFERENCES

- Bond, D.; Devine, P.; Mulligan, P.:** The Accuracy of Imputed Ward and District Codes in the Northern Ireland section of the Central Postcode Directory. A Report for the Department of Economic Development, Northern Ireland Regional Research Laboratory, March 1991.
- Chorley, Lrd.:** In Raper, J.; Rhind, D.; Sheperd, J.; Postcodes The New Geography, pp. x, Longman, London, 1992.
- Department of the Environment,:** Handling geographic information: the report of the Government's Committee of Enquiry. Department of the Environment, London, 1987.
- Gatrell, A.C.:** On the spatial representation and accuracy of address-based data in the United Kingdom. *International Journal of GIS*, 3, 4, pp. 335-48 (1989).
- Haynes, R.:** Inequalities in Health and in Health Service use: Evidence from the General Household Survey. Proceedings of the Fourth International Symposium in Medical Geography, UEA, Norwich 1990.
- Health Departments of Great Britain,:** General Practice in the National health Service. The 1990 Contract. London, HMSO 1989.
- Jarman, B.:** Identification of Underprivileged Areas. *British Medical Journal*, 286, pp 1705-9 (1983).
- Moore, A.J.:** A Geographical Analysis of the Utilisation of General Practitioners in the Coleraine Urban Area, Northern Ireland. Unpublished M.Sc. Thesis, University of Ulster 1988.
- Poole, M.A.:** Religious Residential Segregation in Urban Northern Ireland. In: Boal, F.W.; Douglas, N, H. (ed.), *Integration and Division. Geographical Perspectives on the Northern Ireland Problem.* Academic Press, London 1982.
- Poole, M.A.:** Recent Urban Changes in the Northern Ireland Settlement System. In: Bannon, M.J.; Bourne, L.S. (ed.), *Urbanisation and Urban Development: Recent Trends in a Global Context.* pp.174-

186, Service Industries Research Centre, Department of Regional and Urban Planning, University College, Dublin 1991.

**Raper, J.; Rhind, D.; Sheperd, J.;** Postcodes The New Geography, Longman, London, 1992.

**Senior, M.L.:** Deprivation Payments to GPs: not what the doctor ordered. *Environment and Planning C, Government and Policy* 1991, 9, pp.79-94 (1991).

**Wilson, P.R.; Elliot, D.J.:** An evaluation of the Postcode Addressed File as a Sampling Frame and its use within OPCS. *Journal of the Royal Statistical Society A*, 150, 3, pp.230-40 (1987).