

**STATISTICAL COMMISSION and
ECONOMIC COMMISSION FOR EUROPE**

**STATISTICAL OFFICE OF THE
EUROPEAN COMMUNITIES (EUROSTAT)**

CONFERENCE OF EUROPEAN STATISTICIANS

Joint UNECE-EUROSTAT Work Session on Registers and
Administrative Records for Social and Demographic Statistics
(Geneva, 9-11 December 2002)

Invited paper – SESSION 1

**THE NEIGHBOURHOOD STATISTICS SERVICE
AND THE QUALITY OF ADMINISTRATIVE SOURCES**

Submitted by M. Phillips, D. Brown, M. Cruddas and J. Gastrell,
Office for National Statistics, UK

Introduction

1. The Office for National Statistics is leading the development of Better Information for Small Areas designed to meet the information needs of the National Strategy for Neighbourhood Renewal. This work is part of the Government's response to the problems of deprivation and social exclusion and it is seeking to unlock the rich data held in administrative records by both national and local service providers. A key part of the project is to deal with specific shortcomings and carry out the detailed recommendations set out in the report of Policy Action Team 18 on Better Information¹. This document highlighted a number of problems associated with the provision of information most notably deficiencies associated with access, availability and awareness. In addition, the report underlined the importance of reducing duplication and fragmentation in order to improve the quality of information available for work on neighbourhood renewal including:

- the identification and diagnosis of problems;
- the development of appropriate strategies;
- the targeting of resources;
- the evaluation of outcomes; and
- the measurement of progress.

2. A new internet facility, the Neighbourhood Statistics Service is being set up to meet these needs and to provide a coherent and co-ordinated approach to the provision of information, most notably at small area level. Subsequent sections of this paper will consider the following issues:

- the overall scope of the Neighbourhood Statistics Service;
- key benefits;
- the data development programme and the use of administrative sources;
- quality issues and the use of administrative data; and
- developing a quality framework for the Neighbourhood Statistics Service.

The Overall Scope of the Neighbourhood Statistics Service

3. The Neighbourhood Statistics Service is being developed to help both the public and private sectors tackle problems such as deprivation, social exclusion and crime. The work is being carried out in partnership with many organisations throughout the country, including Central Government departments, the other countries within the UK, local government and other public services.

4. The plan is to draw on the wide range of statistics collected routinely for administrative purposes and, where appropriate, use estimation techniques to overcome gaps and facilitate the use of existing survey data at smaller area levels. When it is fully implemented the Neighbourhood Statistics Service will offer users ready access to a range of social and economic aggregate data on a consistent small area geography. Where exact geographical locations are not confidential, for example, the location of post offices or other services, the aim will be to include precise point references via an underlying geographical information system.

5. As it will take a number of years to reach the goal of full implementation the service is being built up in a series of phases. At present the data relates primarily to wards (NUTS V) or local authority (NUTS IV) areas. However, the inclusion of data from the 2001 Census will make data available at Census Output Area level and later stages will extend the range of data and provide statistics for the smallest area at which quality data can be sustained. When completed the service should make census and other statistics available for thousands of areas across the country.

Key Benefits

6. The aim of this work is to deliver a better technical infrastructure as well as an improved statistical and information resource so that progress towards the high-level policy goals and targets can be measured and tracked over time. The Neighbourhood Statistics Service is the first stage of a 21st Century Small Area Statistics system that will provide easy and ready access to an up-to-date and comprehensive repository of statistical information to facilitate the identification, analysis and monitoring of change for individual areas across the country. The availability of such information will improve the overall development of strategy, the design and delivery of programmes as well as the detailed evaluation of effectiveness. This will in turn improve the allocation of resources, the planning and targeting of action.
7. Access to key statistics will be improved by bringing together information from the 2001 Census and other sources. The availability of comparable statistics at national, regional and local levels will contribute to a wider understanding of the nature and extent of specific problems. Furthermore, it will encourage a wider public user base for official statistics and encourage local participation by individuals and community groups. Local people will be better able to identify and focus on issues that are relevant to them.
8. A common geographic referencing framework will be set up for use by all data suppliers. This framework will be supported by geographic referencing data, accredited methods or specialist software tools in order to facilitate the linking of data records to small areas and harmonisation of geographic standards and classifications for the subscribing datasets.
9. Special measures will be taken to create a more stable geographic environment. In particular, steps will be taken to minimise the impact of the myriad of electoral and administrative boundary changes that occur regularly in the UK, for the derivation of statistics and the development of consistent time series by implementing changes on a common date for statistical purposes.
10. A key aspect of the work will be the development and implementation of stringent disclosure control procedures to protect the confidentiality of personal information and allay any fears that the public may have about the collection and use of local level data.

The Data Development Programme and the Use of Administrative Sources

11. A data development programme has been set up to collect, on a continuous basis, a range of information. This work, guided by recommendations put forward by Policy Action Team 18 and specific requirements identified by the Neighbourhood Renewal Unit, will focus on collecting data for nine key domains and over fifty topics of special interest (see Annex A).
12. At present priority is being given to data sets held by Central Government departments. While some administrative systems already provide the required level of detail, others will need to be developed, extended or improved. In some cases detailed scoping studies may need to be carried out to determine the suitability of specific data sets.
13. Information for the Neighbourhood Statistics Service will be added in tranches as the data development programme proceeds over the next two years. Currently a limited number of data sets are available on the system. Details are summarised in Table 1.

Table 1: Data Sets Available from the Neighbourhood Statistics Service.

Area of Interest	Data Set
Crime	Notifiable offences recorded by the police
Economic Deprivation	Family Credit claimants
	Income Support claimants
Education, Skills and Training	University admissions by place of residence, 1997 and 1998
	Primary school pupils at Key Stage 2, Summer 1998
Health	Attendance Allowance claimants
	Disability Allowance
	Drug mis-users presenting for treatment
	Hospital Episode Statistics, 1999/2000
Housing	Changes of Ownership by Dwelling Price, 2001
	Dwelling Stock by Council Tax Band, 2001
Work	Jobseeker's Allowance claimants
	Incapacity Benefit claimants
	Severe Disablement Allowance claimants
	VAT registered enterprises by industry group, March 2000
	VAT registered enterprises by employment sizeband, March 2000
	Employee jobs, September 1998
	Claimant count - Oxford University ward level estimates
	Claimant count and claimant count rate (revised 29/09/01)
Indices of Deprivation	Indices of Deprivation for wards in England, 2000
	Indices of Deprivation for districts in England, 2000
	Indices of Deprivation for electoral divisions in Wales, 2000
Population and Vital Statistics	Oxford University population estimates for wards in England, mid 1998
	Vital Statistics, 1998
	Resident population estimates for local authorities in England, mid 1998
Other	Parliamentary Electorate 1997 - 98 (England), 1994 -98 (Wales)
Cross-cutting	Countryside Agency's Ward Level Definition of Rural Areas

14. These statistical returns based on aggregate administrative data collected for specific operational purposes are available at various geographical levels ranging from wards (NUTS V) to local authority districts (NUTS IV) as well as specially constituted health authority areas.

15. During 2003 results from the 2001 Census will be disseminated and later phases of the programme will focus on adding data from local government administrative systems and other sources. However, in some cases requirements may have to be met by applying sophisticated statistical techniques or other innovative methods. For example, information from sample surveys may be combined with other data using modelling techniques to produce small area estimates which otherwise would not have been available.

16. It is evident from this outline of the data development programme that administrative records will be used extensively to meet information requirements. However, these data have inherent limitations especially in the United Kingdom where a series of independent systems have been developed in an ad hoc fashion to meet specific needs.

17. Administrative data are collected to support operational and management procedures, satisfy statutory requirements or financial needs. In general, systems are not designed for statistical purposes and as a result information is seen as a by-product rather than a primary output. These circumstances will have far reaching implications for the provision of data and its quality.

18. An awareness of the scope and purpose of each administrative source must be complemented by an understanding of the procedures for collecting the data, processing and analysing the results. In this way it will be possible to identify inherent limitations and determine the best methods for overcoming specific problems. If the Neighbourhood Statistics Service is to meet the needs of policy makers, service deliverers and community workers all information must be robust and fit for purpose.

Quality Issues and the Use of Administrative Data

19. Quality is multi-faceted and difficult to define. No single mechanism is capable of delivering quality; it is based on the combined effects of a range of activities and processes. These complexities are compounded by the varying demands and expectations of the producers and consumers of statistics. Nevertheless, there can be no doubt that the quality of any output is directly related to the quality of key inputs and subsequent processes of verification and validation. In the case of administrative records the strengths and weaknesses associated with the data may be related to a number of inter-related factors including:

- registration practices and recording procedures - the mechanisms for collecting information must be understood. Special attention should be given to whether or not registration is compulsory, the relevance of incentives or penalties;
- quality assurance procedures - methods for checking the completeness of the data must be reviewed alongside processes for correcting inaccurate, incomplete and missing data and for updating individual records;
- continuity - administrative rules, legislative and policy changes at both the national and local level may have implications for the comparability and coherence of the data; and
- data definitions and classifications - variations in administrative procedures or operational constraints may preclude the use of standard definitions and classifications. Any inconsistencies will affect both comparability and coherence.

Developing a Quality Framework for the Neighbourhood Statistics Service

20. Professional standards underpin all work undertaken by the Office for National Statistics and the importance of quality has been enshrined in the recent publication of the National Statistics Code of Practice². This document sets out the basis for developing a quality framework that recognises the need to adopt standard practices and processes for quality measurement and reporting.

21. In accordance with the basic principles set out in the Code of Practice all outputs made available through the Neighbourhood Statistics Service must be 'fit for purpose and of high quality'. More specifically, information must be provided to ensure that users can understand the strengths and weaknesses of the data and make appropriate use of the outputs. At the same time mechanisms are needed to ensure that appropriate procedures are used to assess, quality assure and, if necessary, correct individual data sets.

22. A common understanding of the concept of quality and an appreciation of its complexity is an essential prerequisite for developing a realistic and reliable approach to quality measurement and reporting. In developing a framework for the Neighbourhood Statistics Service the importance of developing consistent indicators has been balanced by the need to take account of the multiplicity of systems and processes associated with the collection, collation, processing and analysis of data from administrative sources.

23. As part of its comprehensive approach to quality the Office for National Statistics has adopted the quality dimensions defined for the European Statistical System. Information provided through the Neighbourhood Statistics Service will be assessed in relation to seven key quality attributes including:

- relevance;
- accuracy;
- timeliness;
- accessibility;
- comparability;
- coherence; and
- completeness.

24. These attributes will be used to develop a more detailed checklist of measures that will be agreed with individual data owners or suppliers. The agreed indicators and reporting specifications will be included in service level agreements. At present work with a limited number of data suppliers has underlined the importance of attributes associated with accuracy, relevance, timeliness, coherence and accessibility.

25. Accuracy is fundamental to quality and the measurement of error is a key indicator. In the case of administrative sources it is likely that identifying and quantifying coverage and processing error will be especially important. However, such measures must be reinforced by an assessment of the overall impact of any errors as well as information about the steps taken to alleviate or minimise specific problems.

26. Standard sets of consistency checks are carried out at the Office for National Statistics on all data sets prior to their acceptance onto the Neighbourhood Statistics website. Comprehensive numeric checks are undertaken to ensure that any anomalies in the data are investigated. Where historic data are available basic trends are analysed and data cross-referenced to ensure that the information is comparable and compatible. Any larger than expected percentage increases or decreases are scrutinised and specific queries are referred back to the data supplier for resolution.

27. Relevance is generally determined in relation to users' needs. In the case of administrative sources, outputs are frequently drawn from systems set up to meet specific statutory requirements or policy initiatives. As a result measurement methods, statistical concepts and returns are linked directly to practical operational or financial requirements. Nevertheless, it will be necessary to take account of the views of ministers, the Neighbourhood Renewal Unit and others, to ensure that the needs of the National Strategy for Neighbourhood Renewal are met.

28. Users want up-to-date information and this is well understood by data suppliers. As a result details of the time lags between the reference point and the date of release together with reasons for any unusual delays are being reported. These basic indicators may be complemented by references to expected improvements in turn-round times as a result of technological developments or operational changes.

29. If the information published by the Neighbourhood Statistics Service is to be used effectively, data definitions and classifications, key statistical concepts and methods must be fully and clearly explained. Although harmonised definitions and classifications will be the recommended standard, operational constraints, financial issues and the burden associated with making radical changes to existing administrative systems must be recognised. As a result the use of alternative definitions or classifications may have to be accepted. In such cases the provision of clear and comprehensive information explaining

the nature and extent of any differences will be essential. Although there is no UK Statistical Law that gives National Statistics right of access to administrative datasets held by other Agencies, a helpful and unique feature of the current work is that significant extra Government funding has been provided to fund the development and collection of data. The availability of funding is particularly aimed at infrastructure development and therefore will help to facilitate the adoption of general standards and methods and improve the chances of harmonising data.

30. A standard template has been developed for statistical, control and technical metadata. Data suppliers are asked to supply comprehensive information on the outputs they produce including details on:

- scope and purpose - a general description of the data including key variables and the main purposes for which the information can be used;
- administrative procedures or methodological issues - includes information on registration or recording practices and procedures;
- concepts and definitions - explaining any discrepancies with agreed standards;
- data classifications - including details on divergence from agreed standards;
- edit and imputation procedures;
- validation and quality assurance - how the quality of the data is checked and the steps taken to validate results;
- geographic referencing;
- disclosure control/confidentiality; and
- sources for further information or assistance.

31. In view of the fundamental importance of geographic referencing and disclosure control priority is being given to working with individual data suppliers to provide relevant advice and support.

32. The importance of providing aggregate statistics for small areas means that addresses must be allocated to the relevant administrative unit, currently electoral wards or local authority areas. All geographic referencing is based on the consistent application of approved sources and methods. At present the Office for National Statistics recommends the use of the 1998 NeSS Gridlink Directory³ and the 1998 Standard Names and Codes (SNAC)⁴. Whereas Gridlink is a look-up directory that relates current postcodes to 1998 administrative areas, SNAC provides a definitive list of names and codes for administrative areas. Data suppliers are required to provide information about the number of matched and unmatched records together with details on areal variations and any changes associated with the amendment or updating of the sources or methods used.

33. The National Statistics Code of Practice is unambiguous about the importance of confidentiality stating that 'no statistics will be produced that would be likely to enable identification of individual characteristics given in confidence'. As publication on the web and access to powerful computers has increased it has become even more important to adopt stringent disclosure controls. In the case of Neighbourhood Statistics these issues are particularly acute because outputs will be produced for areas with low populations and will provide multi-faceted descriptions of each of the constituent areas.

34. In general the risk of disclosure increases as the level of geography decreases. For the Neighbourhood Statistics Service it is anticipated that Census Output Areas with target populations of between 200 and 250 will provide the lowest geography for which outputs will be produced. In addition to the measures adopted by individual data suppliers to protect the confidentiality of their outputs, the Office for National Statistics undertakes a number of additional checks to ensure that information released via the Neighbourhood Statistics Service is non-disclosive.

35. Three case studies have been used to outline the variations in the amount and type of information being provided on quality measures and validation procedures.

Vital Statistics - Deaths

36. All deaths occurring in England and Wales must be registered in accordance with the Births and Deaths Registration Act of 1953. In all cases information is collected by local registrars and simple validation is undertaken at the point of data entry. Further checks for duplicate or missing entries are carried out when the data are received by the Office for National Statistics. Each week around 20 of the 650 registrars are contacted to resolve such queries. As a result 100% coverage is achieved.

37. These simple measures are reinforced during subsequent stages of processing and analysis by frequency and consistency checks on a range of fields as well as the systematic review of overall counts including row, column and other totals. For an annual extract of death occurrences to be acceptably complete it must be taken some months after the end of the data year in order to allow for delays in the registration of deaths from external causes of injury or poisoning. The Office for National Statistics provides a first release ten months after the end of the data year and detailed results two months later. Strict internal confidentiality rules are applied to all deaths data held. Comprehensive advice and support is available to users.

Dwelling Stock by Council Tax Band

38. This data set drawn from Council Tax valuation lists facilitates assessment of the overall number of domestic properties and their distribution across specific council tax bands (a classification relating to the value of the property and therefore correlated to size and available accommodation). Details are given of the numbers of properties allocated to each of the eight standard council tax bands. Explanation of basic concepts is complemented by detailed information on geographic referencing. This indicated that 99.78 % of records were matched to a ward code using Gridlink although more detailed analysis revealed some important geographical variations. A limited amount of information on data quality was provided including basic details on overall accuracy, timeliness, comparability and completeness. This information is not personal and in disaggregated form is publicly available and therefore is not subject to any disclosure control processes.

Small Area Estimation and Quality

39. Small area estimation is the title given to a range of statistical techniques used to produce estimates for small areas where because of limited sample sizes 'standard' survey estimates are unreliable or cannot be calculated. Although all of the quality issues affecting administrative records, surveys and census data apply, there are some additional quality dimensions which merit special attention.

40. The information presented here relates to a particular application currently being developed by methodologists to provide estimates of average household income for small areas (wards). These estimates are currently subject to a quality assurance process and, if approved, will be published next year.

41. The approach adopted for estimation is to derive a relationship between weekly household income as measured by the Family Resource Survey and other sources of covariate information (mainly provided for census and administrative sources) for the wards included in the survey. This relationship is then used to make estimates for all wards.

42. The approach used for estimating income for wards differs from that used to make standard direct survey estimates in that they are model-based and thus depend upon the correct specification of the model, the quality and relevance of the input data sources and the fit of the model. A comprehensive review of the model and the associated estimates is being undertaken and will consider a number of interrelated factors including model validation, accuracy and consistency.

43. As part of the model fitting procedure it is important to ensure that the best possible model is fitted to the data. In order to compare different model specifications and to evaluate their performance several diagnostic tests have been developed. These are designed to evaluate various aspects of the model including bias, goodness of fit, the level of scaling required to calibrate the model based estimates to higher level direct estimates and the predictive power of the models.

44. During the research to develop the models for income a number of different relationships and sources of information were investigated. Having applied the diagnostics and evaluated the fit of the model to the data, the models have been shown to be well specified and the modelling assumptions valid.

45. The main limitation of estimates for small areas either direct from surveys or model-based is that they are subject to variability. Confidence intervals associated with the model-based estimates are produced in order to ensure that the quality of the estimates is clear. However, there are some limitations in terms of consistency.

46. While consistency with the Family Resource Survey published income estimates at the regional and national level is ensured by using standard methods of benchmarking, the model-based estimates will not be consistent with the survey estimates of household income for other geographical areas. Future work may help to overcome this problem.

47. The quality of the results will depend upon the correct specification of the model and the validity of the outputs. Further work is being undertaken to validate the estimates themselves and to compare results with other sources.

Conclusion

48. This paper has outlined the role of administrative data for the development of the small area data and the Neighbourhood Statistics Service and the importance of developing a framework for reporting quality. While recognising the need to comply with the National Statistics Code of Practice and its associated protocols, allowances may have to be made for the special characteristics and inherent limitations of individual data sets. By working co-operatively with data suppliers, expert statisticians and methodologists the Neighbourhood Statistics Service will ensure that the processes, methods and quality of individual returns are reported fully and fairly.

References

- 1 National Strategy for Neighbourhood Renewal, Report of Policy Action Team 18: Better Information, London April 2000.
- 2 National Statistics Code of Practice Statement of Principles, London 2002.
- 3 Gridlink All Fields Postcode Directory, ONS Geography, Office for National Statistics.
- 4 Standard Names and Codes for Administrative and Electoral Geographies in England and Wales, ONS Geography, Office for National Statistics.

Annex A: Neighbourhood Statistics - Domains and Sub-Themes.

<p>Access to services</p> <ul style="list-style-type: none"> ▪ Accessibility of : GP/hospital/other health care Legal advice Leisure facilities Post Offices Schools Shops <p>Community well-being/social environment</p> <ul style="list-style-type: none"> ▪ Caring responsibilities ▪ Participation in community organisations ▪ Perceptions of neighbourhood and service provision ▪ Population turnover ▪ Voting turnout <p>Crime</p> <ul style="list-style-type: none"> ▪ Fear of crime ▪ Numbers of crimes of different types, including domestic burglary, auto crime and violent crime ▪ Offender data <p>Economic deprivation</p> <ul style="list-style-type: none"> ▪ Dynamic measures of low income ▪ Indebtedness ▪ Low income ▪ Wealth/assets <p>Education, skills and training</p> <ul style="list-style-type: none"> ▪ Absenteeism ▪ Adult learning ▪ Driving licences ▪ Early child development ▪ Numbers achieving qualifications ▪ Numbers without qualifications ▪ Pre-school provision ▪ Progress through education ▪ School exclusions ▪ Special needs in schools ▪ Staying on in education ▪ Young people not in education, work or training 	<p>Health</p> <ul style="list-style-type: none"> ▪ Accident and emergency statistics ▪ Disability ▪ Drug and alcohol misuse ▪ Healthy lifestyles ▪ Maternal, infant and early child health ▪ Mortality ▪ Physical and mental health ▪ Social services caseloads ▪ Teenage pregnancies <p>Housing</p> <ul style="list-style-type: none"> ▪ Affordability, including house prices and rents ▪ Composition of dwelling stock ▪ Homelessness ▪ Houses in multiple occupation ▪ Overcrowding ▪ Stock turnover ▪ Unfit housing and disrepair ▪ Vacant properties <p>Physical environment</p> <ul style="list-style-type: none"> ▪ Air quality ▪ Land use, including dereliction ▪ Traffic volume and speed <p>Work deprivation</p> <ul style="list-style-type: none"> ▪ Availability of child care ▪ Dynamic measures of worklessness (duration, persistence and turnover) ▪ Employment ▪ Job losses/notified redundancies ▪ Measures of worklessness ▪ Workless households
--	---