

**A Socio-Spatial Analysis of Perinatal Mortality
in Greater Leicester**

Submitted for the degree of Ph.D.

University of Leicester

1987

Margaret J. Pickton

A Socio-spatial Analysis of Perinatal Mortality in Greater Leicester

Margaret J. Pickton

Abstract

This thesis examines the social and spatial distribution of perinatal mortality in Greater Leicester. The study area comprises the city of Leicester and the surrounding contiguous built up area. Following a discussion of the current status of relevant research in medical geography and perinatal epidemiology the three major data sources used here are described, these being the 1981 Population Census, the Leicestershire Perinatal Mortality Survey to 1982 and the Leicestershire Births records for 1980.

The analysis commences with a classification of census enumeration districts into nine socio-economic 'clusters', reflecting demographic, economic and social differences within the study area. The new small areas created by this classification are then used as a basis for mapping census variables and describing the geography of Greater Leicester.

In a detailed analysis of the perinatal and births data some 21 variables are found to be associated with adverse perinatal outcome. These describe physical, obstetric, socio-economic and health care characteristics of both mother and infant. Both statistical and cartographical methods are used to examine these in a spatial dimension. It is concluded that neither perinatal mortality nor the risk factors associated with it are uniformly distributed with respect to the 'at risk' population.

Further data analysis shows how perinatal, births and census data may usefully be combined. Poisson probability maps are used to show the relative likelihood of perinatal death in different parts of the study area.

The thesis concludes with a summary of the main results and some suggestions regarding directions for future research.

ACKNOWLEDGEMENTS

The production of a thesis inevitably involves a large number of people, to all of these I offer my thanks.

To Professor Michael Clarke and the Department of Community Health I am grateful for the use of the Perinatal Survey data. I alone, however, am responsible for the interpretation and opinions contained in this work.

Too many members of the University's Computer Centre willingly gave their time to solving my problems to mention all by name. Without the tremendous support of these people however, it is certain that parts of this project could not have been completed.

To my colleagues in the Geography Department, both academic and technical, I am grateful for their advice and help. In particular I wish to thank my supervisors, Alan Strachan and David Unwin, who unfailingly offered their time and encouragement.

Finally, I must thank my husband Dave and my daughter Anna. Without the support of the former and the arrival of the latter, in the end I don't think this would have been done.

C O N T E N T S

	Page
Acknowledgements	i
Contents	ii
List of Figures	v
List of Tables	viii
Abbreviations	x
CHAPTER 1: Introduction	
1.1 The thesis in context	1
1.2 Aims of the thesis	2
1.3 Outline of the thesis	4
CHAPTER 2: Medical Geography - Literature Review	
2.1 The meaning of medical geography	7
2.2 Major themes in medical geography	11
2.3 Medical geography in the urban environment	18
2.4 Methods and techniques in medical geography	23
CHAPTER 3: Perinatal Mortality	
3.1 The geographer and studies of perinatal mortality	31
3.2 Perinatal mortality: a summary of publications	43
CHAPTER 4: The Data Sources	
4.1 Introduction	54
4.2 The 1981 Census of Population	54
4.2.1 Introduction	54
4.2.2 The areal basis of the census	56
4.2.3 Using census data	57
4.3 The Leicestershire Perinatal Mortality Survey	59
4.3.1 Context of the survey	59
4.3.2 Choice of method and survey population	65
4.3.3 The questionnaire	65
4.3.4 Published findings	66
4.3.5 Using survey results in a geographical context	67
4.4 The Leicestershire Births Records	68
4.5 Choice of study area	69

CHAPTER 5: Classifying Census Areas

5.1	Introduction	72
5.2	Problems of classifying areal data	73
5.3	Choosing variables for the analysis	76
5.4	A classification of enumeration districts in Greater Leicester - methodology	82
5.5	Conclusion	136

CHAPTER 6: A census map analysis of Greater Leicester

6.1	Introduction	138
6.2	History of Leicester - summary	138
6.3	Univariate descriptions	145
6.3.1	Demographic characteristics	145
6.3.2	Household composition	157
6.3.3	Housing and tenure	168
6.3.4	Socio-economic status	180
6.3.5	Economic activity	190
6.3.6	Other variables	200

CHAPTER 7: Perinatal mortality in Greater Leicester

7.1	Introduction	205
7.2	The Perinatal Mortality Survey - description of the data set	206
7.2.1	The mother's physical characteristics and well- being	207
7.2.2	The mother's obstetric history	211
7.2.3	Socio-economic characteristics	216
7.2.4	Infant characteristics	225
7.2.5	Health care	233
7.2.6	The character of perinatal death in Leicester	240
7.2.7	Summary: Factors of significance in explaining perinatal mortality in Greater Leicester	245
7.3	Leicestershire Births Records - description of the data set	247
7.4	The distribution of perinatal mortality in Greater Leicester	253
7.4.1	Methods for combining data from different sources	253
7.4.2	Allocation of births and mortalities to clusters and census areas - statistical comparison	255
7.4.3	Cartographic comparison of births and mortalities	260
7.5	The distribution of risk	266

CHAPTER 8: Further Data Analysis

8.1	Techniques for combining perinatal and census data: introduction	298
8.2	Some experiments in combining the data sets	300
8.2.1	Principal Components Analysis	300
8.2.2	Correlation coefficient for areal association	314
8.2.3	Risk surface	316
8.2.4	Log-linear methods	325
8.3	Probability Mapping	345

CHAPTER 9: Summary and Conclusions

9.1	Summary	354
9.2	Conclusions	362
9.3	Suggestions for further research	363

APPENDIX 1: Mapping Census Data

A1.1	Introduction	365
A1.2	Choice of class intervals	365
A1.3	Comparing class interval systems	369
A1.4	Examining the effects of data transformation	384
A1.5	Comparing maps of transformed data values	406
A1.6	A recent development in mapping census data - the chi-square statistic	411
A1.7	Conclusion: evaluation and recommendation	417

APPENDIX 2: Computer Programming

A2.1	Introduction	420
A2.2	Programs for data preparation and manipulation	420
A2.2.1	Program MPPOLX	420
A2.2.2	Program MPAGGR	433
A2.2.3	Program MPHOLE	448
A2.2.4	Program MPPTPL	459
A2.2.5	Program MPPTP4	467
A2.2.6	Program MPPTP5	467
A2.3	Programs for data display	474
A2.3.1	Program MPPL0T	474
A2.3.2	Program MPBOX	477
A2.3.3	Control file MPFDW	482

APPENDIX 3: The Perinatal Mortality Survey

A3.1	The Questionnaire	487
A3.2	Modifications to the questionnaire	498

APPENDIX 4: Principal Components Analysis : Results

A4.1	Eigenvalues : First ten components	498A
A4.2	Component Loadings	498B
A4.3	Clusters : Mean values	498C

Journals : List of Abbreviations	499
---	------------

Bibliography	501
---------------------	------------

LIST OF FIGURES	Page
2. 1 Parallel strands in Medical Geography	8
2. 2 Health in an urban environment - factors of influence	20
3. 1 Perinatal mortality: definitions	32
3. 2 A Classification of cause of perinatal death	42
4. 1 National and regional perinatal mortality rates	64
4. 2 Greater Leicester - settlements and communications	70
5. 1 Classifying census enumeration districts - method	88
5. 2 Tukey diagnostics	99
5. 3 Creating new small areas from enumeration district boundaries	121
5. 4 Greater Leicester - 9 clusters	124
5. 5 Cluster 1	127
5. 6 Cluster 2	128
5. 7 Cluster 3	129
5. 8 Cluster 4	130
5. 9 Cluster 5	131
5.10 Cluster 6	133
5.11 Cluster 7	134
5.12 Cluster 8	135
5.13 Cluster 9	137
6. 1 Greater Leicester - council housing	144
6. 2 Greater Leicester - census variables.	
Children aged 0 to 4	146
6. 3 Children aged 5 to 15	148
6. 4 Persons aged 16 to 24	150
6. 5 Persons aged 25 to 44	152
6. 6 Persons aged 45 to retirement	153
6. 7 Greater Leicester - New Small Areas	156
6. 8 Census variables (continued)	
Persons aged from retirement age to 74 years	158
6. 9 Persons aged 75 or over	159
6.10 Persons born in the New Commonwealth or Pakistan	161
6.11 Single person households	162
6.12 Households with 4 or more children	164
6.13 Lone parent households	166
6.14 Mixed households	167
6.15 Households in owner-occupied accommodation	169
6.16 Households in council rented property	170
6.17 Households in privately rented property	171
6.18 Households without exclusive use of bath and w.c.	174
6.19 Households with more than 1.5 persons per room	176
6.20 Households not in self-contained accommodation	178
6.21 Household spaces in purpose-built flats	179
6.22 Professional and managerial headed households (10% data)	182
6.23 Non-manual headed households (10% data)	184
6.24 Skilled manual headed households (10% data)	185
6.25 Semi-skilled manual headed households (10% data)	187
6.26 Unskilled manual headed households (10% data)	188
6.27 Adults with post school qualifications (10% data)	189
6.28 Economically active persons aged 16 or over	191
6.29 Households with 2 or more economically active persons	193
6.30 Economically active persons seeking work	194
6.31 Unemployed people aged 16 to 24	196
6.32 Student population	197
6.33 Married women at work	199

6.34	Self employed persons	201
6.35	One year migrants	202
6.36	Households without a car	204
7. 1	Gestational age: cases and controls	228
7. 2	Change in care during pregnancy and labour: cases	238
7. 3	Change in care during pregnancy and labour: controls	239
7. 4	Greater Leicester: census wards and districts	258
7. 5	Perinatal mortality: distribution of cases	261
7. 6	Distribution of total births: 1980	262
7. 7	Perinatal mortality rate: ratio of survey cases to 1980 total births	263
7. 8	Perinatal mortality rate: chi-square values	265
7. 9	Perinatal mortality: risk characteristics.	
	Mothers of height less than or equal to 150cm.	273
7.10	Zero parity mothers	274
7.11	High parity mothers	275
7.12	Previous adverse pregnancy outcome	276
7.13	Previous fatal pregnancy outcome	277
7.14	Mothers of social class IV or V	278
7.15	Mothers of socio-economic groups 10 or 11	279
7.16	Housewife mothers	280
7.17	Fathers of socio-economic groups 7, 10, 11 or 15	281
7.18	Indian mothers	282
7.19	African and Asian born mothers	283
7.20	West Indian mothers	284
7.21	Hindu mothers	285
7.22	Low birthweight infants (1000 grams or less)	286
7.23	Low birthweight infants (1001g to 1500g)	287
7.24	Low birthweight infants (1501g to 2000g)	288
7.25	Pre-term infants (less than 37 weeks)	289
7.26	Medically induced labour	290
7.27	Breech presentations	291
7.28	Multiple births	292
7.29	Mothers receiving no antenatal care after 20 weeks gestation	293
7.30	Mothers making between 1 and 5 visits for antenatal care after 20 weeks gestation	294
7.31	Mothers receiving more than 20 days in-patient antenatal care	295
7.32	Antenatal clinic session not available at G.P.	296
7.33	Mothers whose GPs are not on the Obstetric List	297
8. 1	Greater Leicester Component maps.	
	Component 1	304
8. 2	Component 2	305
8. 3	Component 3	307
8. 4	Component 4	308
8. 5	Component 5	309
8. 6	Component 6	310
8. 7	Component 7	311
8. 8	Perinatal mortality rate: surface plot by grid square	319
8. 9	Perinatal mortality rate: contour plot	320
8.10	Perinatal risk index: surface plot	322
8.11	Perinatal risk index: contour plot	323
8.12	Standardised perinatal risk index: contour plot	326
8.13	Poisson probabilities: infant population	348
8.14	Poisson probabilities: young children	349
8.15	Poisson probabilities: married women aged 15-44	350
8.16	Poisson probabilities: all births	351

Al.1	Histograms of selected census variables	366
Al. 2	Comparing class interval systems for census variables.	
	Residents aged 45 to retirement - percentiles	371
Al. 3	Residents aged 45 to retirement - arithmetic progression	372
Al. 4	Residents aged 45 to retirement - geometric progression	373
Al. 5	Residents aged 45 to retirement - nested means	374
Al. 6	Residents aged 45 to retirement - break points	375
Al. 7	Residents aged 45 to retirement - mean and standard deviations	376
Al. 8	Households in owner-occupied accommodation - arithmetic progression	378
Al. 9	Households in owner-occupied accommodation - mean and standard deviations	379
Al.10	Households in owner-occupied accommodation - geometric progression	380
Al.11	Households in owner-occupied accommodation - nested means	381
Al.12	Households in owner-occupied accommodation - break points	382
Al.13	Households in owner-occupied accommodation - percentiles	383
Al.14	Households in privately rented property - nested means	385
Al.15	Households in privately rented property - arithmetic progression	386
Al.16	Households in privately rented property - percentiles	387
Al.17	Households in privately rented property - break points	388
Al.18	Households in privately rented property - geometric progression	389
Al.19	Households in privately rented property - mean and standard deviations	390
Al.20	Residents aged 45 to retirement - dispersion diagrams	396
Al.21	Households in owner-occupied accommodation - dispersion diagrams	398
Al.22	Households in privately rented accommodation - dispersion diagrams	400
Al.23	Residents aged 45 to retirement - histograms of transformed data values	402
Al.24	Households in owner-occupied accommodation - histograms of transformed data values	403
Al.25	Households in privately rented property - histograms of transformed data values	404
Al.26	Comparing plotting symbols for census variables.	
	Households in owner-occupied accommodation - proportional squares	407
Al.27	Households in owner-occupied accommodation - proportional triangles	408
Al.28	Households in owner-occupied accommodation - proportional octagons	409
Al.29	Residents aged 45 to retirement - chi-square values	413
Al.30	Households in owner-occupied accommodation - chi-square values	414
Al.31	Households in privately rented property - chi-square values	415

LIST OF TABLES	Page
3. 1 Infant birth and death rates - conventions	35
3. 2 OPCS - publications of relevance in perinatal epidemiology	45
3. 3 Factors associated with adverse perinatal outcome - a summary from the literature	51
4. 1 Using census data - advantages and problems	58
4. 2 Perinatal mortality: national and regional comparisons. Total live births	60
4. 3 Perinatal mortality: national and regional comparisons. Stillbirths	61
4. 4 Perinatal mortality: national and regional comparisons. Deaths under one week	62
4. 5 Perinatal mortality: national and regional comparisons. Perinatal mortality rate	63
5. 1 Census variables used in multivariate classifications	81
5. 2 Classifying enumeration districts - census variables	83
5. 3 Variables used in the cluster analysis	90
5. 4 Classification by iterative relocation: Run 1	92
5. 5 Classification by iterative relocation: Run 2	94
5. 6 Classification by hierarchic fusion: Run 3	95
5. 7 Key diagnostics for 7 clusters derived through iterative relocation from 20 clusters	101
5. 8 Key diagnostics for 7 clusters derived through hierarchic fusion from 30 clusters	106
5. 9 Identifying a 'better' set of clusters - diagnostic statistics	111
5.10 Identifying a 'better' set of clusters - cluster labelling	112
5.11 Reclassifying the enumeration districts in Cluster 1	117
5.12 Key diagnostics for three new clusters	118
5.13 Cluster 1: Variables with a cluster mean greater than the global mean	119
6. 1 Population change in Leicester city	141
6. 2 Greater Leicester: facts and figures	142
6. 3 New small areas with extreme values for census variables	155
6. 4 OPCS and Perinatal Mortality Survey definitions: social classes and socio-economic groups	181
7. 1 Perinatal mortality survey case/control comparisons:	
Maternal age	208
7. 2 Maternal height	209
7. 3 Parity	209
7. 4 Gravidity	212
7. 5 Previous pregnancy outcome: Dead/Alive	212
7. 6 Previous pregnancy outcome: Alive and well/Other	215
7. 7 Mother's social class	215
7. 8 Socio-economic group	219
7. 9 Socio-economic class	220
7.10 Legitimacy	220
7.11 Mother's ethnic group	222
7.12 Mother's region of birth	222
7.13 Mother's religion	224
7.14 Infant birthweight	224
7.15 Gestational age	227
7.16 Type of labour	227
7.17 Presentation at birth	230
7.18 Multiplicity	230

7.19	Attendances for antenatal care after 20 weeks gestation	230
7.20	Number of days in-patient antenatal care	236
7.21	Provision of a clinic session at the G.P.	236
7.22	Initial planned care during pregnancy	241
7.23	G.P's obstetric qualifications	241
7.24	Stage of prgnancy at which death occurred	242
7.25	Was the death avoidable?	242
7.26	Variables associated with perinatal mortality: first summary of results	246
7.27	Leicestershire Births Records: selected variables	
	Birth outcome	248
7.28	Maternal age	248
7.29	Ethnic group	248
7.30	Birthweight	250
7.31	Gestational age	250
7.32	Total births and perinatal deaths: summary of results	252
7.33	Analysis of perinatal mortality by cluster	256
7.34	Analysis of perinatal mortality by census district	256
7.35	Perinatal mortality: summary of risk categories	267
8. 1	Principal components analysis: results	301
8. 2	Comparing the distributions of component scores and perinatal mortality: expected and observed results	313
8. 3	Correlation coefficients for areal association - comparing the distribution of perinatal mortality with census variables	315
8. 4	Calculating a risk index from perinatal data - risk categories	318
8. 5	Calculating a risk index from census data - variables used	324
8. 6	Stepwise logistic regression: variables included in the analysis	333
8. 7	Stepwise logistic regression: summary of results.	
	Model 1: all variables	334
8. 8	Model 2: characteristics of the mother	335
8. 9	Model 3: characteristics of the pregnancy and of the infant	336
8.10	Model 4: health care variables	337
8.11	Model 5: mother's obstetric history	338
8.12	Model 6: socio-geographical characteristics	339
8.13	Models 1 to 6: critical probabilities at the final step	340
8.14	Stepwise logistic regression: assessing combinations of explanatory variables	343
8.15	Stepwise logistic regression: changing the dependant variable	346
9. 1	Additional questions for the Perinatal Mortality Survey	356
A1. 1	Class intervals for map production	370
A1. 2	Assessment of methods of choosing class intervals	391
A1. 3	Residents aged 45 to retirement - descriptive statistics	395
A1. 4	Households in owner-occupied accommodation - descriptive statistics	395
A1. 5	Households in privately rented property - descriptive statistics	395
A1.6	Advantages and disadvantages of the chi-square statistic for mapping	416

SUMMARY OF ABBREVIATIONS

BMDP	Biomedical Programs
CCP	Census Classification Program
ED	Enumeration District
ESS	Error Sum of Squares
G.P.	General Practitioner
ICD	International Classification of Diseases
MAUP	Modifiable Areal Unit Problem
NSA	New Small Area
OPCS	Office of Population Censuses and Surveys
PCA	Principal Components Analysis
PMS	Perinatal Mortality Survey
PNMR	Perinatal Mortality Rate
SD	Standard Deviation
SEG	Socio-economic Group
SASPAC	Small Area Statistics Package
SPSS	Statistical Package for the Social Sciences
UMRCC	University of Manchester Regional Computer Centre

Chapter 1 Introduction

1.1 The thesis in context

Infant and perinatal mortality rates are barometers of the economic and social well-being of societies. Although high mortality rates are characteristic of a developing society, as economic development progresses, standards of medicine and hygiene improve, and social welfare policies are introduced then patterns of health change for the better and as a result death rates decrease. This reduction is not uniform, variations persist at national, regional and sub-regional levels. For example, even among the 'developed' countries of Western Europe there is a three-fold variation in the infant mortality rate (Smith, 1979, quoting 1970 figures). Within England the regional perinatal mortality rate in 1975 varied from 15.8 per 1000 in East Anglia to 21.9 per 1000 in the North West (Adelstein et al., 1980). Locally, rural to urban differentials also exist, in 1973 the county of Leicestershire had a perinatal death rate of 23.2 per 1000 compared with a rate for the city of Leicester of 28.1 per 1000 (see Table 4.5).

The existence of such an exceptionally high perinatal death rate for Leicester city prompted the University of Leicester's Departments of Community Health and Obstetrics to instigate a survey into perinatal mortality in Leicestershire. The results of this survey are used extensively in this thesis.

Following conventional trends in epidemiology, the objectives of the Leicestershire Perinatal Mortality Survey were to identify, quantify and assess the factors contributing to the perinatal death rate in the county (MacVicar et al., 1977). Data were obtained by means of a comprehensive questionnaire, eliciting information from case notes and personal interviews. Many of the

factors investigated had already been identified by other surveys as having explanatory significance (see Chapter 3), examples included variables such as maternal age, social class, smoking habits and obstetric history. Other variables, such as ethnicity, diet and obstetric practice were felt to be of particular relevance in the Leicestershire context. The survey itself was therefore well placed to offer both supporting evidence for previous work, and new insights into particular perinatal problems.

To the geographer, the availability of the survey data offered several exciting possibilities. First, it provided a comprehensive and detailed picture of the socio-medical environment of the survey population. That individual data had been collected was a particular advantage. It meant that not only could inferences be drawn for other individuals, but also that comparisons could be made with data from other sources, such as the population census. Secondly, although for obvious confidentiality reasons, names and addresses were omitted from the survey data, each case and control was given a postcode as a spatial reference, from this it was possible to derive a coordinate value which in turn allowed each observation to be allocated to a specified areal unit.

The process by which these areal units were identified, were allocated with perinatal mortalities and were described by perinatal 'risk factors' is a major part of the present research. Other foci are outlined in the research objectives listed below, and the approach followed is summarised in the subsequent thesis plan.

1.2 Aims of the thesis

There are two broad sets of objectives for this research, those related to the data and concerned with the new knowledge

which may be derived from these, and those related to the methods used and the applicability of these to various research problems.

Data objectives:

- i) To describe the social, economic, demographic and health characteristics of Greater Leicester.
- ii) To establish 'risk factors' comprising those characteristics most highly associated with perinatal death.
- iii) To identify, using these risk factors, those sections of the population most likely to be 'at risk' of a perinatal mortality.
- iv) To locate areas of greatest risk within the study area based on the distribution of the 'at risk' population.
- v) To compare the distribution of mortalities with the areas of risk identified.
- vi) To determine that part of the perinatal mortality distribution which may be ascribed to factors which are geographical in nature and responsive to ameliorative action, and to isolate these from any other medical, social, cultural and economic factors.
- vii) To provide information that would be of assistance to policy makers and implementers in improving perinatal health care.

Methodological objectives:

- i) To investigate the benefits to be derived from combining major data sets from different sources.
- ii) To make use of available computer technology to develop a fully automatic system of extracting, manipulating and displaying data.
- iii) To examine the validity of the graphical approach in the analysis of socio-medical problems, and in particular to

investigate the usefulness of the computer-drawn map.

1.3 Outline of the thesis

This research places emphasis on the social and spatial dimensions of perinatal mortality and presents a geographical analysis of the health experience of an urban population. Chapter 2 places the present work within the context of current research in medical geography. Given the scarcity of geographical research into either perinatal or infant mortality, two secondary themes are expanded. First, aspects of health and health care in other urban areas are considered, and secondly, an attempt is made to describe and assess the wide range of methodologies currently employed within the field of medical geography.

Chapter 3 reviews the large body of mainly epidemiological research into perinatal mortality. Having firstly defined some of the terms most commonly used in perinatal studies, it then extends the methodological section of chapter 2 by examining the problems inherent in the geographer's use and interpretation of perinatal statistics. The second part of the chapter summarises the findings of previous national and regional perinatal surveys, and it concludes with a list of factors found to be associated with adverse perinatal outcome.

Chapter 4 describes the three main data sets used in this project. These were the 1981 Population Census, the Leicestershire Perinatal Mortality Survey and the Leicestershire Births Records. The choice of study area is explained and its limits defined. Particular attention is given to the use of the different data sets within the current research.

Chapter 5 is the first of five further chapters concerned with the manipulation and analysis of the major data sets. This chapter describes in detail the process by which variables were

extracted from the 1981 census, transformed through multivariate analysis into component scores, and then used as the basis for a classification of the 847 enumeration districts in the study area. The resulting nine 'clusters' are then mapped and their dominant characteristics are described. This provides a reasoned analytical description of the study area, and a framework for subsequent analysis.

Chapter 6 extends the description of Greater Leicester firstly by providing as background information an outline of the recent history of the city, and secondly, by mapping and commenting on each of the census variables in turn. These univariate descriptions apply to aggregations of the population - in some cases individual enumeration districts, in others, groups of enumeration districts (or 'new small areas') which have been combined because of their spatial contiguity and common cluster membership.

In contrast, the data sets described in **Chapter 7** pertain to individuals. Selected data from the perinatal survey are examined with respect to their possible influence on perinatal outcome. Thus, the search is made for associations between physical and obstetric, maternal and infant, socio-economic and health care characteristics and perinatal death. The analysis is based on a comparison between the 'case' and 'control' populations. A summary is given of those factors found to be significantly associated with adverse perinatal outcome. The second part of chapter 7 is devoted to an analysis of the Leicestershire births data. These data are used both to extend the findings of the perinatal survey, and more particularly, to assess the validity of the survey's 'control' population in representing the 'normal' pregnancy and birth. The chapter concludes with a description of the distribution of risk,

particularly the distribution of mothers exhibiting 'risk characteristics' as identified earlier in the chapter. Individual cases and controls are plotted and the resulting patterns are compared with those of births and perinatal deaths in total.

In Chapter 8 further analysis of the major data sets is undertaken. The main objectives are to compare and contrast the area-based census data and the individual based survey and births data. Having outlined some of the general problems likely to be encountered, the chapter describes and discusses the results of several data manipulation exercises. These include a principal components analysis, the calculation of correlation coefficients to describe areal association, the calculation and plotting of a perinatal 'risk surface' based on both survey data and some new census variables, the application of log-linear methods in assessing the relative importance of different 'explanatory' factors, and, finally, the calculation and plotting of Poisson probability maps.

The concluding chapter of the thesis draws together the findings from each of the preceding chapters, showing how the results of each stage of research are used as a foundation for the next. Reasons are given for the various decisions taken throughout the project, and suggestions made of avenues for profitable future research.

Several appendices are included in this thesis. These include notes on some of the many computer programs written for the project and on the choice of style and class interval selection for the computer drawn maps. A copy of the original perinatal survey questionnaire is provided.

Chapter 2. Medical Geography - Literature Review

2.1 The meaning of medical geography

Medical geography is essentially a 20th century discipline having achieved a degree of prominence since the second world war. Related work, however, can be traced as far back as the 4th century B.C. and the Hippocratic school of medicine (McGlashan, 1972b) or, more recently, to the 19th century medical topographers such as Hennon (1830) and Powell (1846) and medical climatologists such as Clark (1829) and Shapter (1849). The concern of these authors was to relate disease and ill-health to the physical environment, a theme which continues to the present day.

Modern medical geography encompasses a diversity of philosophies, methods and subjects of interest (see Figure 2.1). In very general terms it has evolved from a primarily deterministic stance to an approach in which behaviouralism and social relevance have key roles to play. This chapter reviews some of the main strands of this work, especially as they are relevant to the current research.

Several authors (Shannon and Dever, 1974; Learmonth, 1978; Pyle, 1979) have drawn attention to the existence of two 'medical geographies'. Howe (1972, p.2) provides a definition of the first:

"Medical geography may be defined as the comparative study of the incidence of disease and the distribution of physiological traits in people belonging to different communities throughout the world and the correlation of these data with features of the environment".

This is commonly referred to as 'disease ecology', 'geographical epidemiology' or 'geographic pathology' (Mayer, 1982b). The second medical geography is more usually known as the geography of health care delivery. Phillips (1981, p.2) describes its main themes as

Figure 2.1. Parallel strands in Medical Geography

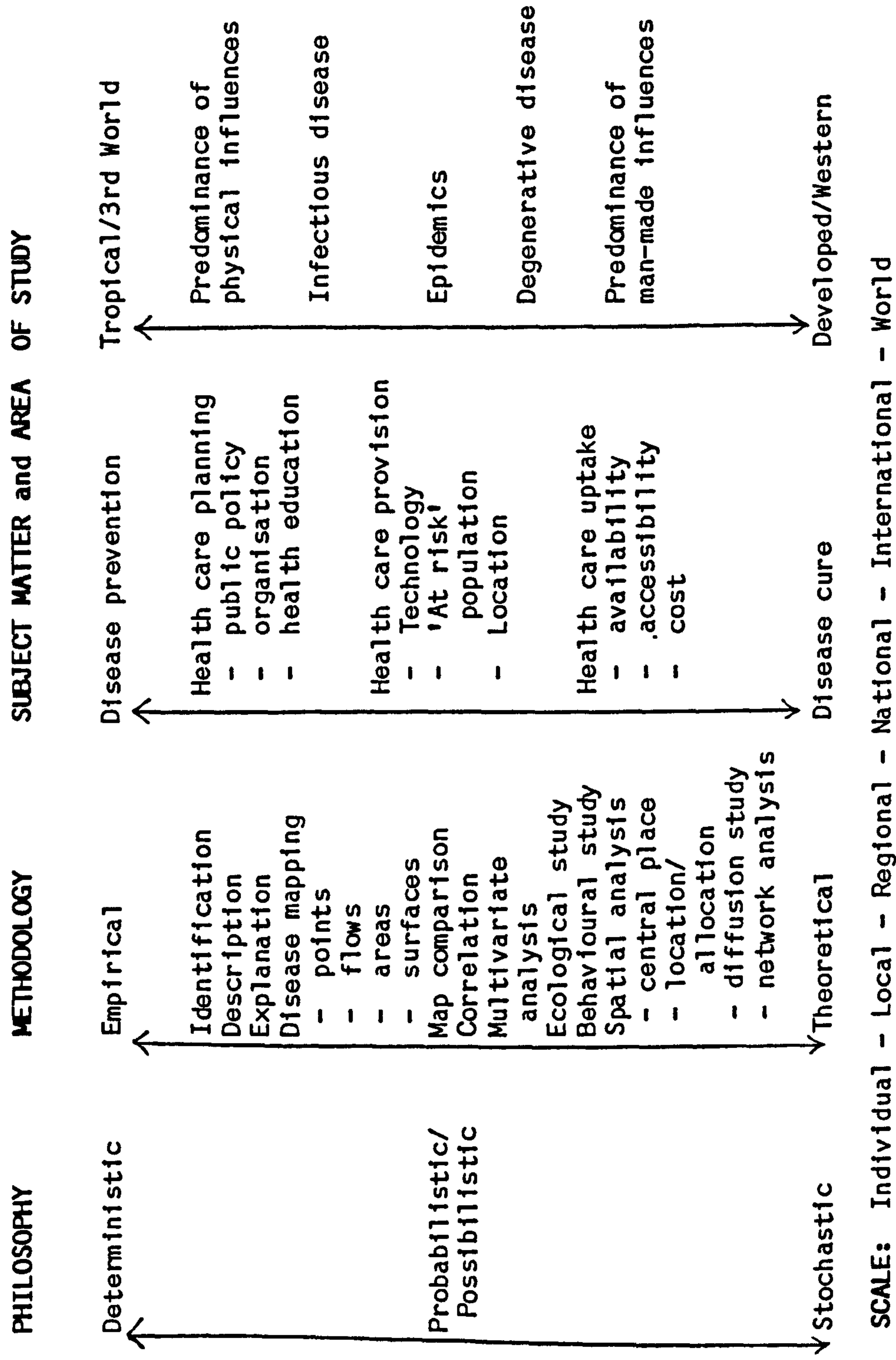
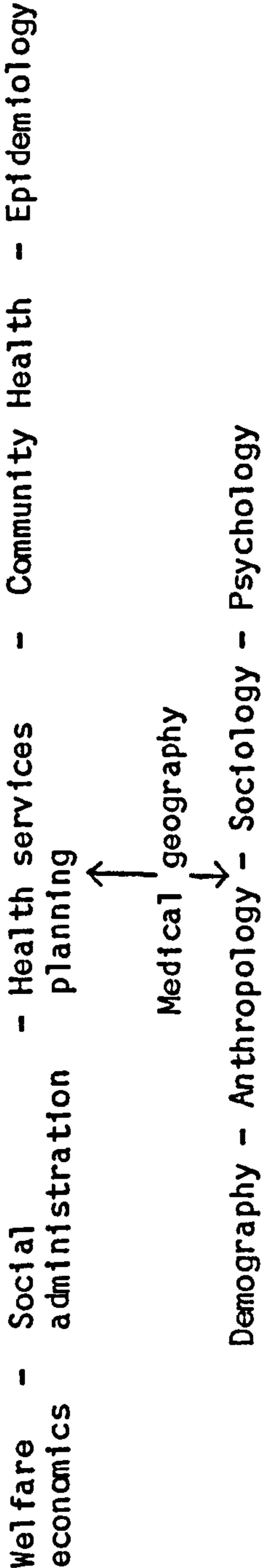


Figure 2.1 (Cont.)

INTERACTION WITH OTHER DISCIPLINES:



"the spatial analysis of health care, health behaviour and the provision of medical services".

The distinction between, and synthesis of, the two approaches has been the focus of much recent interest. Giggs (1979) offers perhaps the most useful attempt at reconciling the apparently divergent interests. He suggests that a simple health system model should include three elements: the first is concerned with the spatial patterning of ill-health and mortality, the second with patterns of the physical and human environments (and particularly those characteristics which might be detrimental to man's health), and the third with the spatial pattern and use of the health care delivery system. The first element may therefore be viewed as the identification and description of a problem condition, the second offers explanation, and the third ameliorative action. All three are necessary for an effective contribution to the health of society.

Mayer (1982b, p.224) furthers this argument, suggesting

"an initial and essential link between geographical epidemiology and spatial planning of health services is therefore that adequate performance of the latter presupposes knowledge of the former".

Knowledge of disease distribution, dynamics and spread, and of the distribution of the population at risk, provides the predictive ability which is the strength of a unified approach.

Other authors have sought reconciliation between the different approaches via their subject matter and scale of analysis. Learmonth (1978) views the ecological method as a contribution to the prevention of disease (exemplified by the emphasis on 'cause and effect' in ecological studies) whilst health care research, with its interest in welfare, deprivation, accessibility to care and so forth, is concerned with the cure. The two branches are therefore directly relevant to one another:

"The ecological side of medical geography must always be important because it may play its part, with sister sciences, in

social, community, or preventive medicine. Prevention is better than cure, and cheaper, but if cure there must be, then the rational planning of health services is vital" (Learmonth, 1978, p.239).

Pyle (1979) describes continuity between the two approaches with respect to scale of analysis. Although examples of both types of research may be found at all scales, he notes that the main thrust of disease ecological research has been at the international, national and regional scales, whilst health care research, frequently behavioural in emphasis, has been predominantly regional, intra-urban and at the individual level. He accounts for this variation by constraints on the availability of data.

This author believes that to be overly concerned with the 'split' in the discipline is to divert attention from the main issue - the betterment of human health. Within medical geography there will probably always be conflicts in philosophy, method and subject matter, but the very eclecticism of the discipline may be seen as a virtue. In the same way as geographers as a whole have traditionally adopted concepts and techniques from other disciplines, so too can medical geographers, bringing these together in a holistic manner and as a result contributing new thoughts and ideas to the solution of health problems.

The remainder of this chapter will describe some of the more influential research of medical geographers, this will be followed by a summary of the recent literature most relevant to this thesis, and finally some examples will be given of the methods and techniques most commonly used in medical geography.

2.2 Major themes in medical geography

Dr John Snow's 'On the mode of communication of cholera' is regarded as a 'classic' essay on the subject of medical geography. Prompted by the outbreak of cholera in various English cities, and

in London in particular, Snow set out to determine the extent of a 'cause and effect' relationship in the spread of cholera. In the second edition of his work published in 1855, Snow included a map showing the distribution of deaths from cholera in an area about 300 yards square, centred on a water pump in Broad Street, central London. During a period of only 10 days in 1854 five hundred deaths were recorded within this 'cholera field'. Snow suggested that the incidence of cholera was only among people drinking from this pump and demanded that its handle should be removed immediately. This demand was met, and with the removal of the handle came a marked drop in the number of new cases of cholera.

Snow's work clearly relates the natural environment (in his case, water) to the spread of disease. This was the direct approach taken by many of the early writers, and is apparent in work at all spatial scales (see, for example, Petermann, 1852; Hirsch, 1886; McKinley, 1935; Simmons, 1944).

The French medical geographer Jacques M. May is frequently credited with spearheading the resurgence of interest in medical geography in the post war period. In addition to publication of a series of world scale disease maps, May wrote extensively on his understanding of medical geography. He viewed the subject as a study of the relationships between pathological and geographical factors (May, 1977). The five 'pathological' factors he identified were as follows:

- (i) Causative agents such as viruses and bacteria
- (ii) Vectors spreading these causative agents to man, for example flies and mosquitoes
- (iii) Intermediate hosts which are essential to the vectors, for example, rats
- (iv) Reservoirs acting as hosts to the infection, storing it until it is picked up by man

(v) Man himself

His 'geographical' factors comprised the physical/environmental, human and social, and biological. May applied this thesis to his discussion of disease and the environment, emphasising two groups of disease - those resulting from parasitic infection and those related to the consumption of specific foods.

In the 1960s British geographers, most notably Howe and Learmonth, were responsible for major advances in the discipline. Learmonth's main interests lay in the medical geography of India and Pakistan and his regional studies of these areas were published widely (Learmonth 1958, 1961, 1964, 1967). His 1958 study of survival, mortality and disease in Indo-Pakistan has been described as a milestone in the development of medical geography (Blunden, 1983).

Howe's main contribution of this period was his 1963 'National Atlas of Disease Mortality in the United Kingdom'. Howe devised the 'standardised mortality ratio' (SMR) as a means of allowing for local variations in population age structure, thereby facilitating nationwide comparisons of ill-health. He recognised the problems in calculating epidemiological indices and sought to achieve statistical significance with his SMR (Howe, 1970b).

Developments in disease ecological studies and disease mapping were paralleled by increasing interest in matters relating to health care provision, and in particular, concern for the inequitable distribution of health care resources. In Britain, despite the underlying principle of the National Health Service:

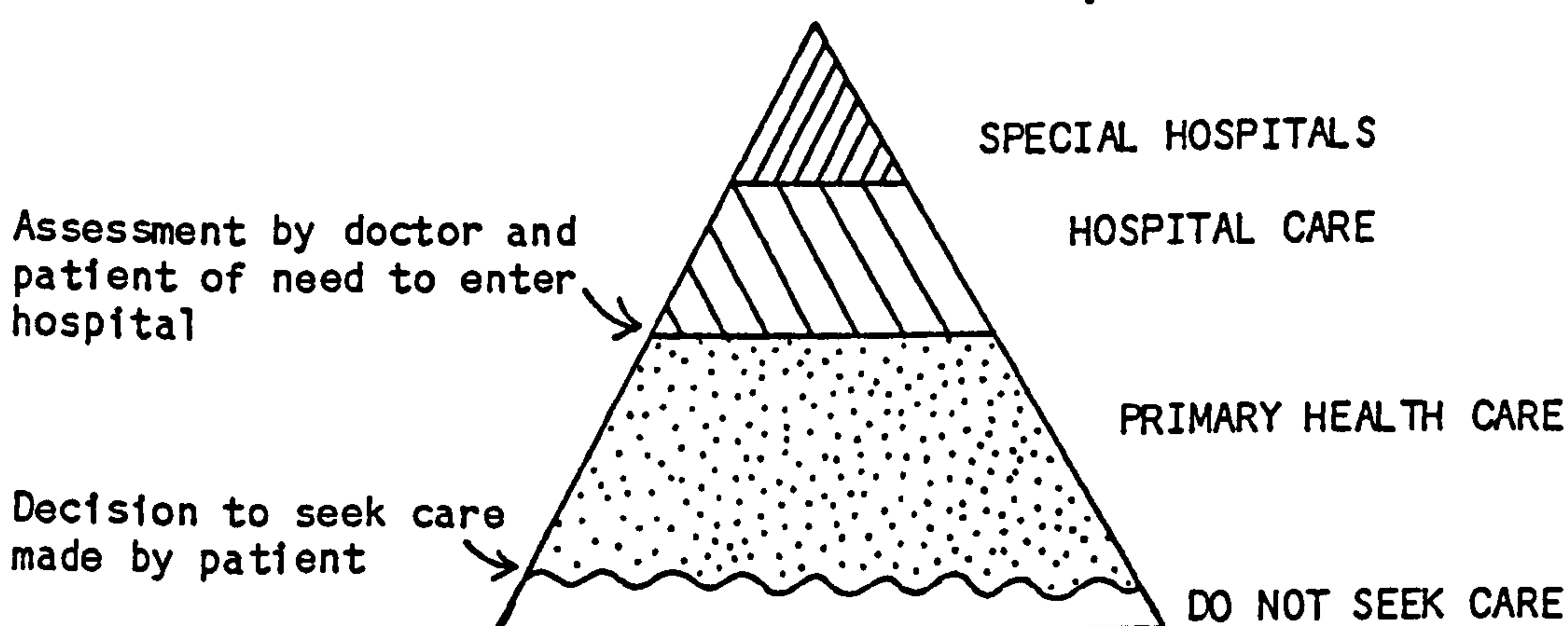
"....that medical care should be available to all who need it, with no discrimination on grounds of income, race or creed"
(Sumner, 1971, p.46)

there were still felt to be unacceptable inequalities within the health care system. Hart exemplified this view in his 'inverse

law of care', that is

"the availability of good medical care tends to vary inversely with the need of the population served".
(Hart, 1971, p412).

Studies of both provision and utilisation of health care proliferated in the 1970s, approaches ranging from the highly theoretical 'location-allocation' type modelling to the essentially empirical behavioural research. Internationally, the wide range of health care systems (Roemer (1977) suggests five types: free enterprise, welfare state, transitional state, underdeveloped and socialist state) prompted a diversity of research of equal prominence. McGlashan's (1977) concept of the medical care hierarchy was universally applicable. McGlashan noted that different medical conditions required varying levels of service, which due to financial constraints would have to be selectively allocated. Uptake of service he illustrated by means of a 'morbidity triangle', thus:



Source: McGlashan (1977, p773)

McGlashan suggested further that the triangle could be divided vertically to allow for the different types of health care system.

Fundamental to research in medical geography is the notion of space. Location, accessibility, distance (actual and perceived) and related concepts form the core of research and the major contribution geographers can make to health care planning

and medical science. Whether the pattern reflects public policy or personal decision-making, inequalities in the spatial distribution of all levels of health care are recognised. For example, in the provision of primary care Knox (1979b) suggests several factors which, assuming they are not constrained by public policy, might influence doctors' locational decisions:

- (i) the distribution of potential patients, especially those in the higher income groups
- (ii) personal ties of home, medical school etc.
- (iii) the availability of working space - hospital beds, offices and so forth
- (iv) the proximity of other GPs, to provide professional interaction and locum services
- (v) the presence of low income groups and of ethnic and racial minorities
- (vi) the attractiveness of the physical environment.

The result of these 'pull' and 'push' forces is an imbalance between medical needs and resources. This is exemplified by pre-NHS Britain in which Knox identified both 'doctor-poor' regions (for example the North East, Lancashire and the Midlands) and sub-regions (particularly the inner areas of conurbations). Although largely negated by policy controls and incentives in the 1950s, subsequent inflation and financial pressure has allowed the re-introduction of regional disparities.

Spatial factors feature strongly in studies concerned with uptake of medical services. Shannon et al. (1969) provide an early review of this literature by examining 'the concept of distance as a factor in accessibility and utilization of health care'. They classify studies into those taking an 'area approach', in which health personnel or facilities are related to the size of the area's population, and those which look at the relationship

between individual patient and provider using some measure of distance for comparison. Gravity models, central place theory and migration studies provide a theoretical basis to the latter. Shannon and Dever (1974) give examples of this work.

Researchers concerned with disease and patterns of ill-health have also made increasing use of more rigorous statistical methodologies. The most significant developments have been in the use of 'associative analysis' and diffusion theory.

'Associative analyses' assume greatest importance within medical geography where direct causal relationships cannot be found, since they serve to highlight ranges of variables which may be of importance in explaining the aetiology of specific diseases. This is especially the case in consideration of Howe's 'diseases of civilisation' (1977). Cardiovascular disease, cancers and mental illness, for example, are known to have certain 'risk factors' associated with them which might also have spatial associations with other distributions such as patterns of physical, social or economic factors.

Giggs (1973) studied the distribution of schizophrenics in Nottingham and found that most first admissions to hospital were from addresses in or near the city centre. He then used factor analysis on 29 variables (mainly drawn from the population census) to identify a 'social/environmental schizophrenia dimension'. Giggs concluded that social and environmental factors combined in Nottingham to form a 'milieu' in which schizophrenia was more likely to develop. Low social status, high unemployment and low social cohesion were all found to be significant. Since these unfavourable traits were of greatest intensity in the inner, slum areas of the city it followed that these places were where one might expect to find a high incidence of the disease. In Giggs's

words:

"Here, as in other large cities, there are pathogenic areas which seem to destroy mental health" (Giggs 1973, p.71).

There are numerous other examples of this type of work in the literature. For example Rose (1976) examined the epidemiological effects of the urban environment; Briggs and Leonard (1977) the association with the disadvantaged population and Levy and Herzog (1977) the effects of crowding. Similar work by MacGlashan (1982a) examined the concept of the epidemiologic transition, Benjamin (1982) looked at regional differences and Taylor et al. (1985) the aetiology of diarrhoea. Each of these authors attempts a multi-factorial explanation of health patterns.

The study of infectious diseases and of disease epidemics lend themselves to the application of diffusion theory. Based on Hagerstrand's early work on innovation diffusion (Hagerstrand, 1967) it is implicit that diseases will spread outwards from a central point "like ripples from a stone thrown into a pond" (MacGlashan 1972b, p.12). This outward spread has both a temporal and spatial dimension. Brownlea (1972) developed relatively sophisticated models in an attempt to understand the diffusion of infectious hepatitis in Wollongong, Australia. He described the spatial spread of an epidemic wave originating in Wollongong which travelled about 15 miles north and 8 miles south over a three to five year period. Both the initial outbreak and subsequent 'nests' of infection were explained by physical environmental factors, notably sewerage and drainage conditions. Ecological factors such as community age structure, rate of residential growth, distance from infection nodes, levels of immunity, extent of water pollution and linkages in the settlement pattern were noted. Principles of diffusion theory have also been applied to the study of measles (Haggett, 1972), cholera (Stock, 1976) and

influenza (Hunter and Young, 1971).

2.3 Medical geography in the urban environment

This thesis is concerned with the distribution of perinatal mortality across an urban area. To the best of this author's knowledge there has been no directly comparable research within the field of medical geography. Epidemiologists have examined the factors contributing to perinatal death in great detail, and some of their findings are described in the next chapter, but the spatial component and hence the geography of the problem has largely been ignored. Even studies of infant mortality by other social scientists (such as Willie, 1959; Shah and Abbey, 1971; Jibbu, 1972; and Wilson, 1973) have either failed to make explicit the spatial element, or, like Howe (1970a), have examined the problem at a much larger scale. The literature reviewed in this section therefore, is selected on the basis of its general relevance to the present work. Social and spatial variations in other aspects of urban health are considered, in particular inequalities in child health, ethnic health and patterns of primary health care.

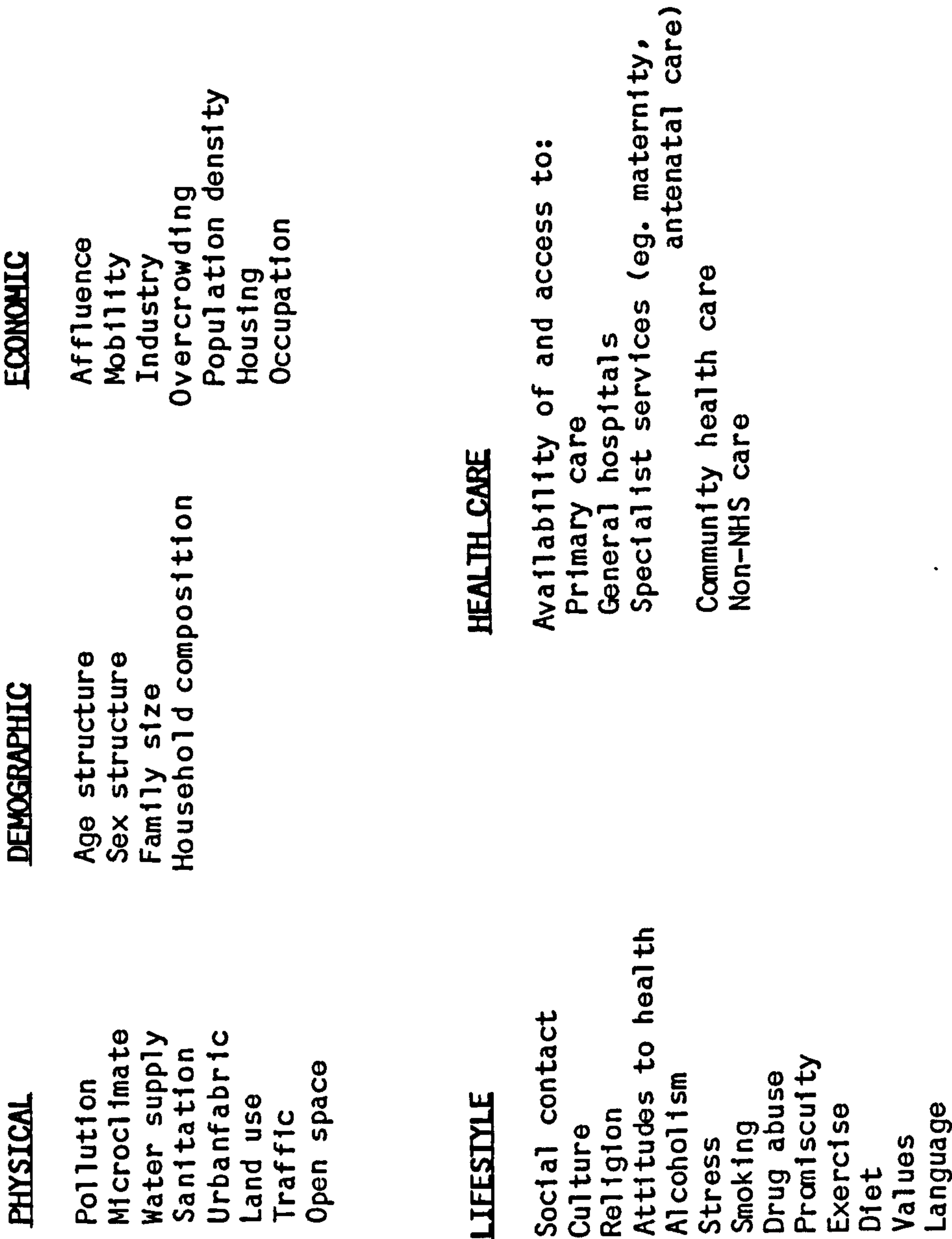
Giggs (1982) has reviewed the recent literature on intra-urban variations in health patterns. He identifies four of the main areas of research described above - disease mapping, ecological associative analysis, disease diffusion studies and the geography of health care - and for each approach gives examples of recent work. Importantly, he also considers some of the methodological and conceptual problems associated with these studies, these will be considered in more detail in section 2.4.

In studying health patterns in urban areas several important factors have to be considered. On the one hand there are physical, social, demographic, economic and lifestyle factors, all of which initially contribute to patterns of disease and ill-

health, while on the other, the planning, provision and uptake of health care facilities contribute to the eventual outcome (see Figure 2.2). Some authors (e.g. Weinstein, 1980) suggest that the urban environment may be pathogenic (able in itself to cause or produce disease), others suggest only that there are associations between urban characteristics and certain disease types. For example Carstairs (1981) examined associations between health state and multiple deprivation in Glasgow and Edinburgh. She concluded that for most measures of health there was evidence of greater mortality and morbidity in areas of greater deprivation. Interestingly, this result did not hold true for either perinatal or infant mortality, although she suggests this might be due to the low event counts involved. Donaldson (1976) focussed on urban/suburban differentials in health. By selecting some 40 census enumeration districts within Teesside to represent the inner core, the 'good' suburbs and council estates, he compared standardised mortality ratios for various conditions (including bronchitis, emphysema, suicides, accidents and violence) and concluded that the inner area suffered the worst health experiences. Of particular interest here were the higher neonatal mortality rates he found within the inner area. These he explained by social and environmental conditions, illegitimacy, maternal age and several indications of poorer child health (such as less regular attendance at child health clinics, high incidence of chest problems, measles and head infestations, and variations in age-related weights and heights).

Kirby and Scott-Samuel (1981) have considered aspects of both health and health care in the inner city. With regard to health they note that inner areas are frequently populated by people less able to cope with the poorer quality of living

Figure 2.2 Health in an urban environment - factors of influence



typically found there, for example the elderly, single parents, the mentally ill, the poor and the unskilled, and that as a result concentrations of high mortality are to be expected, especially for diseases such as lung cancer, heart disease, chronic bronchitis, pneumonia, accidental poisoning, suicide and rheumatic heart disease. Aspects of health care provision are conflicting. On the one hand, they note that the **quantity** of provision in inner areas is above average, due to factors such as the centralization of regional facilities, externalities such as public pressure against locating mental health clinics in suburban areas and spatial inertia; on the other, they suggest that the **quality** of provision is poor. In particular the provision of primary health care (for most people the first, if not the only contact with the medical services) is characterised in the inner city by the single doctor surgery, the ageing GP and the redirection of funds to the major hospitals at the expense of primary and community care. Despite the advantage of physical proximity they suggest that accessibility to care may still be poor. Low levels of car ownership and the often elderly age structure of the population account for this. Donaldson (1976) suggests further that cultural barriers may make care facilities as inaccessible as geographical barriers.

Given the multi-cultural nature of Leicester's population, variations in the health experiences of different ethnic groups in other areas are of especial interest. Considering first their experiences of health care, Rathwell (1984) suggests that ethnic minorities have different 'needs' and demands of the health services. He lists several factors which he feels should be accounted for in the provision of care. These are cultural differences and attitudes towards care, a possible distrust of doctors and hospitals, the stigma attached to some diseases,

lifestyle differences (such as diet and religion), language difficulties, the attitudes of health care providers, and finally, the additional social and psychological stresses encountered by the immigrant population. Johnson and Cross (1983) argue a different case, that ethnic minorities should not be seen as problem groups and that they do not place excessive demand on NHS facilities. These authors explain that, within their study area of the West Midlands where it occurs, the higher usage of care is linked to needs related to broader sociological inequalities and not to unjustified demand on the part of certain ethnic groups. They counter the alternative criticism, that Asians 'opt out' of the NHS (thereby creating a 'reservoir of ill-health') with evidence of improved uptake of immunisation services among this population. Pearson (1983) summarises the prevailing attitude toward health care for ethnic minorities that despite evidence of racial discrimination and disadvantage, studies have resulted in little action. Minority cultures are viewed as pathological and pathogenic and public policy which assumes social integration as a pre-condition for better health is ineffective.

Several authors have also found disparities in the quality of health among ethnic minorities. Sometimes this is related directly to population migration and the subsequent environmental change and lifestyle modification (Bandaranayake, 1983), in other examples poorer health among ethnic minorities is attributed to social pressures from the white majority population (Donovan, 1983), to physiological variations among ethnic groups (Clarke and Clayton, 1983), to inequalities in the standard of care either available to, or taken advantage of, by certain groups (Clarke and Clayton, 1983) or to hazards of the immediate living environment forced on an immigrant population without access to large

financial resources (Bandaranayake, 1983). Although ethnic minority populations have been shown to have the worst experience of most types of ill-health, for example tuberculosis, measles, typhoid, dysenteric illnesses and perinatal and infant mortality, there are some conditions which are less common among such groups, such as most cancers and chronic bronchitis.

2.4 Methods and techniques in medical geography

A summary was provided in Table 2.1 of the methodologies most frequently employed by medical geographers. In the main, both the methods used and the problems encountered are typical of geography as a whole, thus problems of scale and aggregation, ecological fallacies, statistical significance, data compatibility, and so on are important in medical geography. The purpose of this section is to focus attention on the application of some of these methods, and the particular problems which arise, in a medical geographical study. For further information the reader is referred to any of the general reviews of medical geography, for example Pyle (1979) or McGlashan and Blunden (1983), or to the specialist reviews of techniques of which Mayer (1982a), Stimson (1983) and Gesler (1985) are recent examples.

The starting point of most research is to identify a problem condition, to describe its distribution and if possible to offer an explanation. The earliest studies were limited to just these activities. Disease mapping is a logical extension to a verbal description. In increasing levels of sophistication, maps may comprise patterns of points, lines (or flows) areas and surfaces. They may be used as both descriptive and analytical tools. McGlashan and Harington (1976) analyse deaths from cancer among South African gold miners using various types of map. First they plot absolute case numbers (one dot equals one death). Secondly they convert the data into a rate map which allows for population

spread and is portrayed by a choropleth and, thirdly, they assess the significance of the variation in cancer deaths by comparing the data with a Poisson distribution. Finally, they extend this still further by combining the concepts of gradient and probability mapping and plotting 'iso-mells', or lines of equal probability of cancer. In the process, the function of the maps is expanded from simple illustration to detailed spatial analysis.

The production of disease maps has recently been influenced by technological developments. As long ago as 1969 Howe recognised the value of computer maps as research tools; subsequent improvements in the speed, efficiency and reliability of computers have greatly enhanced this value. Bickmore and Stocking (1982) have produced a series of computer-drawn maps as part of the WHO Onchocerciasis Control Programme in the Volta River Basin. In their view

"the philosophy that underlies computer cartography assumes that the data base is more important than any map that derives from it" (p.24).

The collation of data from various sources is therefore central to their research. Their 'cartographic data base' comprises both a base map marked with physical features, disease control areas and international frontiers; and various data collected from epidemiological surveys of the area. The advantages of such a data base (or 'geographic information system') lie in its flexibility, especially with regard to which information is included on a map, the ability to update information on an ongoing basis, and the facility of combining these data with those from other surveys or pertaining to other diseases.

Discounting for the time being the normal problems associated with the presentation of data the major problem from the medical geography perspective is deciding which measure of

morbidity or mortality to display. No simple epidemiological index completely characterises the impact of a disease on a community (Howe, 1979). Thus measures of incidence, prevalence, case fatality and even standardised mortality ratios must carefully be interpreted in the light of the research objectives and constraints.

Data problems comprise probably the largest set of difficulties the medical geographer has to overcome. To begin with medical data are frequently confidential and availability is limited. Secondly, measures of morbidity are notoriously difficult to obtain, yet many diseases which have a great social impact do not actually cause death (King, 1979). Even mortality data may be of dubious accuracy. For example the diagnosis of cause of death as given on death certificates was found to be unreliable by Fox (1972), as was the location of the deceased's place of residence. Aggregation, in order to obtain significant results, increases data stability but this is often at the expense of information (Stimson, 1983). Moreover, it also leads to problems of non-uniqueness due to the possibility of differing aggregations giving differing results. The process of standardisation and the choice of appropriate ratio values introduce further inaccuracies. These are compounded by problems which arise when comparing or combining data sets from non-matching scales or time periods. Other problems are peculiar to certain diseases (eg. notifiable diseases) or to specific methods of data collection (eg. sampling strategies, response rates, interviewer bias, error and under-reporting).

Following on from disease mapping, various authors have experimented with methods of map comparison. These range from the simplest visual comparison of maps, to the increasingly complicated Lorenz curves, correlation coefficients, coefficients

of areal correspondence, and difference map approaches (Gesler, 1985). The most commonly used statistic is the **correlation coefficient** - normally either Pearson's product moment or Spearman's rank correlation statistic. Dependant and independant variables may be compared individually, or, in the case of the ecological correlation, collectively. It is frequently the case that variations in individual demographic or socio-economic variables are too weak to be associated with disease patterns, but following some multivariate manipulation (say a factor or cluster analysis) can produce a significant correlation.

It is interesting that medical professionals such as Scott-Samuel (1977), Carstairs (1981) and Maclean (1982) have appreciated only relatively recently the value of **social area analysis** in community health studies. Scott-Samuel uses both census and health data to classify EDs in Liverpool into 5 'families' (or clusters), then searches for relationships between these and health conditions (infant mortality, triple vaccination and infectious disease). He suggests there are five broad areas of application of social area analysis within community medicine:

- (i) as an information base for resource allocation and health services planning
- (ii) to highlight health care needs in small areas, independantly of health district boundaries
- (iii) to identify sample area types with known socio-economic characteristics for subsequent comparison with mortality and morbidity measures, and for the evaluation of innovations
- (iv) to identify socio-economic factors in the aetiology of disease
- (v) to identify 'target groups' for health education.

Geographers have, of course, been undertaking similar

'ecological' studies for some time (eg. Purola et al. (1968), Fox (1972), Giggs (1973), Briggs and Leonard (1977) and others).

In common with other branches of geography some authors have used a behavioural approach to the study of health problems. In a few cases these methods have been used to investigate the aetiology of disease. Examples include Craig's (1982) investigation of the transmission of diarrhoeal diseases in Bangladesh and Taylor et al.'s (1985) study of the same disease in Grenada. More frequently, principles of behavioural theory have been applied to health care uptake. The strength of the behavioural approach, which relies heavily on personal and detailed interviews, lies in its ability to establish individual attitudes and perceptions of health and health care. For example, in 26 interviews with Afro-Caribbeans and Asians in London, Donovan (1983) discovered several factors to be important to the respondents, namely religion, work, racism, traditional remedies and diet. Wood and Barr (1982) interviewed both providers (G.P.s) and consumers of primary health care in Greater Manchester to establish patterns of both provision and use of the health services. Although they reported difficulties in interpreting some of their results (for example an experiment in measuring G.P. catchment areas by asking each G.P. to draw his practice area on a map showed their perceptions of size to be far greater than expected and this resulted in considerable overlap of areas) the authors were able to conclude that there was "little systematic difference between doctors practicing in different parts of the study area" (Wood and Barr, 1982, p.2).

Studies of patient behaviour are beset by problems of a conceptual nature. Stimson (1983) suggests that it is essential to have a theoretical base in behavioural terms before examining consumer behaviour. Many authors (eg. Gross 1972) propose complex

models for investigating consumer behaviour, in which the use of health services is determined by 'enabling' factors, 'predisposing' factors, accessibility, perceived health level, and environmental factors. The problems arise when attempts are made to quantify these. Assessment of both the demand and the need for health services (generally considered to be separate issues) is as difficult to determine as it is to measure accessibility.

Much research into health care provision is essentially normative. The assumption is made that it is desirable to optimise some measure of equity, access or efficiency (Mayer, 1982b) and highly theoretical spatial models are built to describe optimal location patterns. One broad group, the location-allocation models, is widely used in the planning of both private and public facilities (see for example Beaumont, 1981)

"Simply stated, the rationale underlying location-allocation problems is optimising (maximising or minimising) the value of an objective function subject to a set of specified constraints. The fundamental problem is to optimise the location of a set of p supply facilities (CH's) in order to minimise the aggregate transport costs associated with a set of n demand points. The problem is uncapacitated in that there is no maximum limit on the number of patients at any particular CH. Thus, at the optimal solution, it is assumed that each demand point is allocated to the nearest facility." (Sixsmith, 1982)

There are clearly many drawbacks to such models since they depend for their success on the accurate definition of demand for the health service and on the assessment of transport costs. For some facilities, such as community hospitals for the elderly, mentally ill and infirm, it is even doubtful whether transport costs should be considered important at all. These units rarely provide emergency services and their patients receive relatively few visitors.

Although as geographers the main focus of most work is on the spatial distribution of health facilities, the role of time, particularly as a factor in travel cost and accessibility, is also

considered important. Weil and Weil (1980) have examined the role of time in various aspects of health care - for example the response time of emergency services (this is especially important in cases of cardiac arrest), patient travel time to the nearest appropriate facility, and the use of medical services within overall activity patterns.

Mayhew (1981) has gone a step further by calculating a 'velocity field' which applies to emergency medical centres in large cities. Because of the variation in traffic congestion in such cities Mayhew suggests that travel time is more important than distance so he has created a set of 'isochrones' to delimit points of equal accessibility.

Time plays an equally important role in studies of disease diffusion. Knox (1971) argues that in order to examine the spread of disease in space, one must also have an understanding of the frequency of disease in terms of time. Many infectious diseases (for example influenza) occur in time series cycles, with epidemics of varying proportions occurring at almost regular intervals (Pyle, 1979). Examples of the literature on disease diffusion are to be found in Haggett (1972) and Girt (1978).

This summary of the main methodologies practiced by medical geographers is by no means exhaustive. Mention must also be made of research which has adopted less common techniques, for example Jones' (1982) work in historical medical geography, Schoeman and Mahajan's (1977) use of the Delphi method in assessing community health needs, Joseph's (1985) discussion of 'accessibility' indices, and Lovett et al.'s (1985) recent use of a Poisson regression model for analysing small numbers of deaths.

The most striking feature of a review of research done by medical geographers is the breadth and flexibility of their

approach. They show a willingness to adopt and adapt methods and techniques from other branches of geography and from other disciplines which has and will foster the rapid development of the discipline.

Chapter 3. Perinatal mortality

3.1 The geographer and studies of perinatal mortality.

Despite the increased interest in medical geography over recent years, few geographers have addressed the problem of infant mortality and none that of perinatal mortality. The lack of suitable data, particularly at appropriate scales or units of analysis, must be the main reason for this, since the potential contribution geographers can make to perinatal studies is enormous. The perinatal death rate is seen by many to be an indicator of social and economic well-being from local to international level; it has been shown to be both socially and spatially discriminant; and importantly, it can be ameliorated with appropriate health care provision and economic aid.

Unfortunately, the collection and use of perinatal data present the geographer with many problems. First of these is the **definition of terms**, a summary of which is provided in Figure 3.1. It is usual for 'perinatal' deaths to include both 'stillbirths' and 'early neonatal' deaths. As used in the Leicestershire Perinatal survey, these are defined as follows:

A stillborn child is

"a child which has issued forth from its mother after the 28th week of pregnancy and which did not at any time after being completely expelled from its mother, breathe or show any other sign of life"

(Births and Deaths Registration Act 1953, amended by Population (Statistics) Act 1960, and quoted by MacVicar et al., 1977, p.31)

An early neonatal death is normally classified according to age

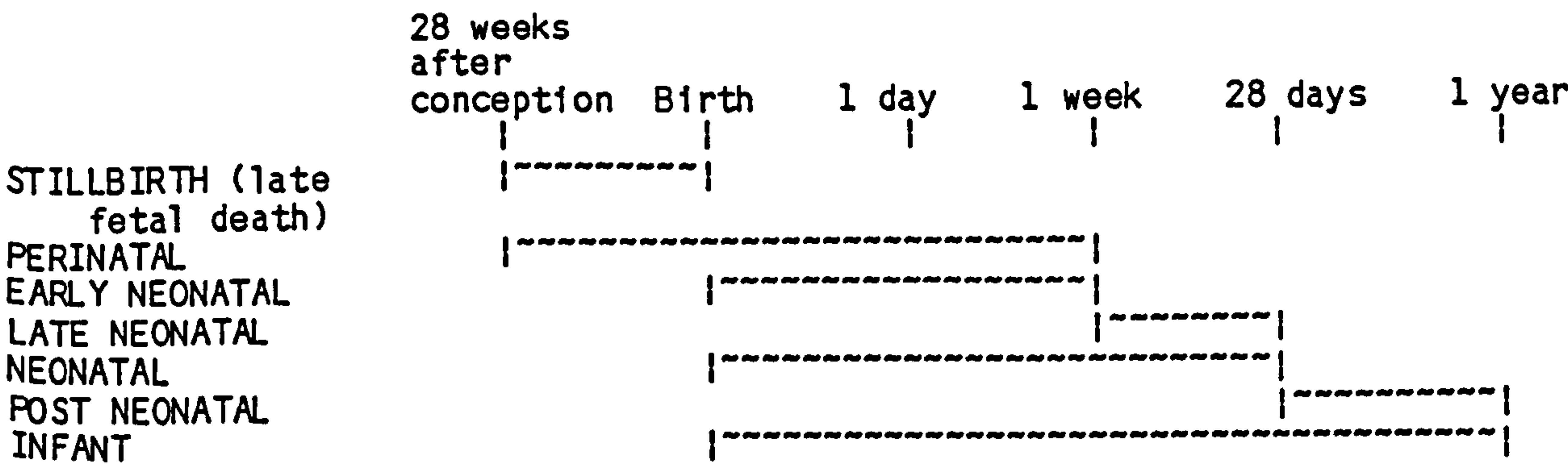
"by single days for the first week of life (under 24 hours, 1,2,3,4,5, 6 days)"

(Manual of the International Statistical Classification of Diseases, Injuries and Causes of Death (1967,p.471) quoted by MacVicar et al. 1977, p.31)

This artificial delimitation of the perinatal period avoids the sometimes critical problem of establishing exactly when a baby died, specifically before, during or immediately after birth, but

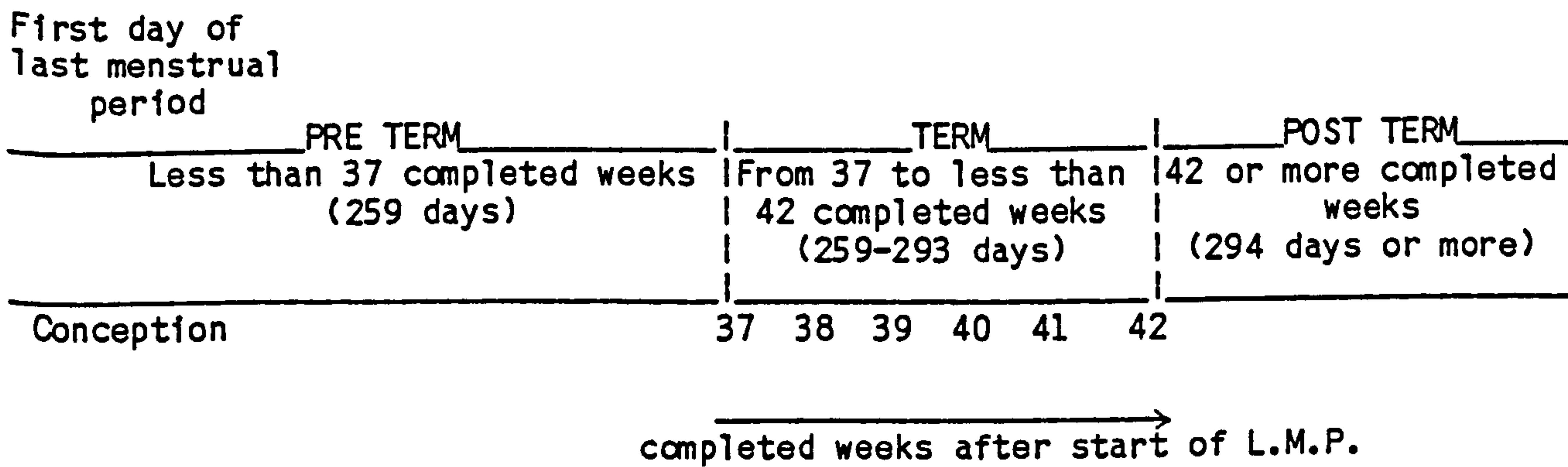
Figure 3.1 Perinatal Mortality : definitions

(i) Infant mortality



Source: after Macfarlane (1979b) and Lambert (1976)

(ii) Gestational age



Source: Macfarlane (1979b p.128)

(iii) Perinatal statistics: ICD recommendations

	International statistics	National statistics
Minimum birthweight	1000g	500g
Minimum gestational age	28 weeks	22 weeks
Minimum body length	35 cm	25 cm

Source: Macfarlane (1979b)

may nevertheless cause subsequent difficulties in interpretation. For example, Chalmers (1979a) points out that the risk of death to the fetus or newborn child changes rapidly from the time of conception on. In fact

"within 6 months of conception nearly half of all conceptuses will die. Yet by the time the perinatal period is reached the overall chances of survival are high: between 95% and 99% of those human beings alive at the beginning of it are alive at the end of it" (Chalmers 1979a, p.1063)

This has implications for the use of perinatal data since clearly a small change in the boundaries used to define the perinatal period can have a substantial effect on the estimate of risk of death. Using gestational age as the lower cut off point requires the mother to be sure of her dates, yet it is known that this is often not the case, especially in pregnancies which end unsuccessfully. At the upper cut-off point medical skill and technology can influence the length of time a baby may be kept alive, even though, for example, it may suffer from a condition which is incompatible with independent existence (Chalmers, 1979a).

In addition to these difficulties, the baby which is born prematurely and alive, but which subsequently dies presents a different problem of classification. The International Classification of Diseases defines a live birth somewhat loosely, i.e.

"Live birth is the complete expulsion or extraction from its mother of a product of conception, irrespective of the duration of the pregnancy, which after such separation, breathes or shows any evidence of life, such as beating of heart, pulsation of the umbilical muscles, whether or not the umbilical cord has been cut, or the placenta is attached; each product of such a birth is considered live born" (Quoted by Macfarlane, 1979b, p.129).

Thus, the baby born prior to 28 weeks gestation if born dead would be unclassified, but if born alive only to die within a few days would become a neonatal mortality. Similarly, classification of the baby which dies 'in utero' prior to 28 weeks but is delivered as a stillbirth after this date is also ambiguous.

Use of the perinatal statistics, once classified, depends on some standardisation by the host population. In common with much social scientific research, this frequently entails the calculation of appropriate rates or ratios. Table 3.1 lists numerators and denominators as conventionally used in the literature and unless otherwise stated, as followed here.

Further standardisation may be undertaken, for example to allow the effect of specific variables to be taken into account. Mallet and Knox (1978) calculated 'Standardised Perinatal Mortality Ratios' (SPNMRs) to allow for variations in birthweight distribution across Area Health Authorities. This showed that an area's relatively high perinatal mortality rate did not necessarily reflect a poor standard of obstetric care, but was rather related to regional differences in birthweights.

Nortman (1974) performed a similar exercise in investigating the effect of maternal age on pregnancy outcome, while Naeye et al. (1973) contrasted weight and body measurements of 'normal' with stillborn infants to examine the effect of maternal nutrition on fetal growth.

Identification of a 'case' population and a set of 'controls' against which to standardise form the basis of a commonly used method of perinatal survey, the 'case-control' survey. The main objective of this type of survey is

"to compare a series of patients with a given disease with those of a control group who did not have the disease, in terms of the absence or presence of the hypothesized etiologic or risk factors (Clarke and Clayton, 1981, p.636)

It is important that the quantity and quality of information should be the same for both groups, and if controls are chosen on the basis of shared characteristics with the cases then care must be taken that the usefulness of key variables is not eliminated by this process. For example, the Leicestershire Perinatal Survey

Table 3.1 Infant birth and death rates - conventions

Numerator	Denominator
Live births	per 1000 home population
Illegitimate live births	per 1000 live births
Stillbirths	per 1000 total births (ie live + stillbirths)
Perinatal deaths	per 1000 total births
Early neonatal deaths	per 1000 live births
Late neonatal deaths	per 1000 live births
Neonatal deaths	per 1000 live births
Post neonatal deaths	per 1000 live births
Infant deaths	per 1000 live births

Source: Macfarlane (1979) and Lambert (1976).

'matches' cases and controls on the basis of 'intended place of delivery', this of course results in a one-to-one case-control ratio for this variable, rendering it useless in terms of any future analysis and an unfortunate choice from the geographer's point of view.

Other methods of data collection for perinatal studies include the 'retrospective' and the 'longitudinal' survey. To a limited extent the case-control study makes use of retrospective data in that some important and necessary information pertains to the mother's medical and obstetric history, thus the difficulties associated with generating a data set from such secondary sources must be addressed. Problems include the availability and quality of relevant information, (Clarke and Clayton (1980) note the poor quality of routine medical records as a source of information for identifying risk factors); differences in recording and recall; the continuity and comparability of information derived from varying sources; the reliability (or 'reproducibility') of such information and its validity (or 'accuracy'). Arguably, an ongoing case-control comparison minimises these potential disadvantages by incorporating the same bias, if any, in both populations.

A longitudinal survey, being one in which a population or sample (such as 'all pregnancies') is examined over an extended period of time, is not considered feasible for a perinatal survey. A conventional longitudinal survey aims to establish three sets of facts: the rate of new occurrence of a disease (ie. its incidence); its natural history and outcome; and the association between initial characteristics and the risk of future disease occurrence. Thus the greatest return will be achieved in the study of a disease reaching epidemic proportions, not in that of a relatively low incidence event such as perinatal mortality where it would be necessary to collect data from 98 pregnancies for

every 2 which result in a perinatal death.

Whichever method of data collection is employed consideration must be given to the interpretation of resultant statistics. Macfarlane (1980) notes that variations in mortality rates among small areas could arise because of several factors, namely

- "1. Chance variation
 2. Differences in the completeness of data
 3. Differences in the population and its environment
 4. Differences in the quality of medical care available to the population."
- (Macfarlane, 1980, p.274)

Explanations of apparent relationships may be difficult to substantiate from the given data. For example, variables not included in the immediate data set may account for some or all of an association; Friedman (1980) gives an example of this. An association is found between parental loss (ie. one or both parents being dead) and low back pain; this is a spurious relationship which may be explained by a third underlying variable to which both of these are related, namely age. It is often possible to prevent these spurious associations from arising by knowing of previous findings and if necessary controlling for the missing variables, for example by examining the low back pain/parental loss association for specific age groups. This of course is not a problem which is unique to medical geography, it arises whenever inference is to be drawn from sample data and can usually be avoided by reasoned interpretation.

Whether the association can be applied to other populations is a test of both the representativeness of the current sample and the success of the interpretation. In some cases adjustment of the results must be performed in order to make the sample reflect the total population from which it is drawn. Clarke and Clayton (1980 and 1981) described in some detail how they have adjusted their

figures to allow for the bias they found in the Leicestershire control sample.

The problem of causality in interpreting perinatal statistics is considered by Chalmers (1979b). He suggests that five criteria must be satisfied before a causal link can be established. These are as follows:

- "1. The time sequence in the relationship must be correct
 2. There must be consistency of association when the observations are replicated
 3. The relationship must be specific, that is, there should be no exceptions
 4. The strength of the relationship should be consistent with a causal link
 5. There should be a coherent explanation of the presumed causal nature of the association in the light of current knowledge"
- (Chalmers, 1979b, p.189)

It is clearly difficult to identify a true causal relationship from observational data, leading many authors to suggest the presence of 'associations' rather than causes and effects. Even associations can apparently be contradictory, as shown by Chambers (1978) and Chalmers (1979a). Both note that whilst cross-sectional analyses have found that the risk of stillbirth increases with decreasing inter-pregnancy interval and with increasing parity in multigravidae, this causal relationship is not supported by longitudinal surveys examining overall reproductive history. These explain that, on average,

"women with a history of pregnancy loss will have both a large total number of pregnancies and shorter intervals between pregnancies"

(Chalmers, 1979a, p.1064)

This phenomenon is known as 'reproductive compensation' and should be allowed for in the interpretation of the association between parity, birth interval and perinatal death. Chambers (1978) notes the potential danger of allocating apparently normal individuals such as high parity mothers with a history of successful pregnancies, to 'high risk' status on the basis of associations

which are not fully understood.

A final problem area in perinatal studies is the identification of the cause of death. The geographer is interested in the cause of death insofar as the different causes have different aetiologies and will be responsive to different courses of action. Unfortunately the disease which eventually results in the baby's death may be the final outcome of a complex series of events, any or all of which may have contributed to the eventual demise. Barson (1980) gives a typical example:

"A sixteen year-old unmarried primagravida from social class V has an antepartum haemorrhage at 28 weeks gestation followed by spontaneous labour and the premature delivery of an infant who rapidly develops respiratory distress syndrome and dies on the second postnatal day. The post-mortem examination demonstrates hyaline membrane disease and intraventricular haemorrhage".
(Barson, 1980, p.228)

The author then illustrates how each medical specialist (paediatrician, obstetrician, pathologist and epidemiologist) will prefer to identify a single cause of death, rather than accept the more accurate multifactorial explanation.

To quote Barson again,

"a cause of death, like beauty, lies only in the eye of the beholder"
(Barson, 1980, p.229)

Different authors have different ways of approaching this problem. In the Leicestershire survey both 'direct' and 'antecedant' causes of death are recorded ie. the direct cause of death is that which eventually determines the outcome, the antecedant cause is that which may be considered as the start of the causal chain of events. (MacVicar et al., 1977). The OPCS death certificate on the other hand, requests the whole sequence of diseases in reverse order (ie. the disease which started the sequence is shown last) as well as a list of diseases which contribute to the main cause but are not part of the fatal

sequence (Adelstein, 1980).

A second problem in identifying cause of death applies particularly to longer term studies. Chalmers and Macfarlane (1980) argue that variations in the incidence of disease may appear to occur either for genuine reasons, or because the likelihood of spotting the disease changes. Clearly this has implications for the identification of secular trends.

A final problem in an examination by cause of death concerns the statistical analysis of essentially rare events. Statistical significance may be impossible to achieve without some form of aggregation or classification, both of which must be acceptable on medical and statistical grounds.

Various authors have attempted to produce classifications of the causes of perinatal deaths. So far, the most widely used of these are derived from the work of Baird et al. (1954) in Aberdeen. Walker (1980) describes the seven categories of cause of death, namely deformity; antepartum haemorrhage; premature cause unknown; toxæmia; mature cause unknown; maternal disease and trauma. This classification however, and others based on it, have a disadvantage for a geographical study in that they are based on clinical instead of wider environmental causes and they have scant regard for common aetiology.

Baird and Thomson (1969) and Baird (1976) have subsequently improved the classification by dividing causes into two groups: 'environmental' and 'obstetrical'. The former is taken to include unexplained deaths of premature babies, congenital malformations and diseases in the mother, while the latter comprises problems such as toxæmia, pre-eclampsia, rhesus incompatibility and some mechanical complications of labour. The authors see environmental causes as being related to the mother's social class and so forth and hence, in their view, unavoidable, but the obstetrical causes

they consider could be often prevented by a high standard of obstetric care.

Wigglesworth (1980) has produced the first classification of cause of death which explicitly recognises the importance of common class treatment and therefore implicitly the significance of similar aetologies. Ideally he suggests that groups should be "as far as possible mutually exclusive, easy to recognise, and carry implications for perinatal care" (Wigglesworth, 1980, p.685)

To have implications for clinical management it is essential that all members of his groups should each respond to the same sequence of actions. Thus all his class labelled 'congenital malformations' may be detected early in pregnancy by prenatal screening; likewise, 'asphyxial conditions developing in labour' may be controlled by obstetric management during this period. Wigglesworth suggests appropriate responses for each of the five classes he identifies (the other three are 'normally formed macerated stillbirths', 'conditions associated with immaturity' and 'other' specific conditions).

Clarke (1982) has developed Wigglesworth's classification still further by identifying the four major subgroups shown in Figure 3.2. An advantage of his fairly broad 'action'-oriented approach is that it avoids the problem of classifying deaths to a single cause; moreover, by adopting this classification over a period of time it is possible to evaluate the successfulness of clinical methods such as screening programmes, and even the effectiveness of obstetric care in different institutions (although the latter must be undertaken with due regard to factors which are not explicitly included, such as birthweight).

Of Clarke's four classes, 'asphyxia in labour', which covers asphyxial conditions developing in labour and all types of birth

Figure 3.2 A Classification of cause of perinatal death

STILLBIRTHS	EARLY NEONATAL DEATHS
CONGENITAL	MALFORMATIONS
ASPHYXIA	IN LABOUR
Fresh stillbirth (Any weight)	Term infants (38/42 +)
MACERATED STILLBIRTH (Remainder)	IMMATURE (Remainder)

Source: Clarke (1982), Fig 1., p.105

trauma, is the only group that may be unrelated to social and environmental factors - instead it is greatly influenced by obstetric management. The others might all raise questions about the maternal environment (in its widest sense) as well as the provision of health care, both of which the geographer is well equipped to examine.

In conclusion therefore, the geographer has both the tools and the experience to consider not only **who** suffers ill-health in perinatal terms, but also **where** and possibly **why**. This information can most importantly be used for planning the distribution of maternity services to ensure that those in greatest need of care have optimal access to it.

3.2 Perinatal mortality : a summary of publications

In examining the factors which contribute to a perinatal outcome the geographer is performing a similar role to that of the perinatal epidemiologist and it is within the field of **epidemiology** that much relevant work is to be found. Epidemiology is defined as follows:

"the study of disease occurrence in human populations".
(Friedman, 1980, p.1)

or,

"the study of the distribution of a disease or condition in a population and of those factors which influence their distribution"
(Lilienfeld, 1958, p.649)

Like the social geographer, the epidemiologist is concerned with a process - be it simple or multi-factorial - that leads to a specified outcome, such as urban decay or perinatal death. The objects of study are in both cases groups and the methods employed have much in common. For example,

"...one of the central concerns of epidemiology is to find and enumerate appropriate denominators in order to describe and to compare groups in a meaningful and useful way"
(Friedman, 1980, p.9),

a statement which might equally apply to a geographical study.

The perinatal epidemiologist is specifically concerned with the epidemiology of morbidity and mortality in the perinatal period. As Chambers (1978) observes, perinatal medicine has certain advantages for the epidemiological approach; namely, the potential for satisfactory data collection, the availability of an easily-defined population, and unambiguous (if crude) measures of outcome.

In addition to specially commissioned surveys (discussed below) much epidemiological information is readily available from public sources. These include returns from death certificates and disease notifications, occupational mortality statistics, the Hospital In-Patient Enquiry, Hospital Activity Analyses, and in-patient statistics from the Mental Health Enquiry. Table 3.2 provides a summary of the most relevant information. A detailed discussion is given by Barker and Rose (1979). The Population Census and the General Household Survey provide more general data.

Independant surveys vary widely and include both 'ad hoc' surveys, ranging from simple descriptive surveys based on small groups of cases to major cross-sectional population surveys, and more extensive and longer term longitudinal studies. The nature of the survey will reflect its purpose, with such aspects as size (both geographical and numerical), time scale, type of population studied (sex, age, health status etc.), sampling method, examination technique (e.g. questionnaire design) and so forth, being constrained by factors such as cost, availability of historical data (necessary for any type of retrospective study), and both long and short term research requirements.

The results of epidemiological studies may include the description of geographical distributions; the identification of secular trends or cyclic changes; the association of certain individual characteristics with disease outcomes; and in

Table 3.2 OPCS - publications of relevance in perinatal epidemiology

Series	Topic
FM1	Birth statistics. Live and stillbirths. Legitimacy, seasonality, age of parents, parity, multiplicity, area of residence, place of confinement, birthplace of parents, social class, fertility, abortion.
DH1	Mortality statistics. Deaths by cause, age and sex, standardised mortality ratios (including infant mortality).
DH2	Mortality statistics - cause. Age and sex breakdown (including under 1 year) by cause of death as classified by the ICD.
DH3	Mortality statistics - childhood and maternity. Stillbirths and infant deaths. Legitimacy, sex, cause of death, place of confinement, maternal age, parity, length of gestation, birthweight, area of residence and maternal death. Accompanying DH3 monitors deal specifically with infant and perinatal mortality.
MB3	Congenital malformation statistics - notifications.
MB4	Hospital in-patient Enquiry. Maternity statistics are published in the form of monitors. Length of postnatal stay and births by region of residence are included.

Further analyses of the above data are found in other OPCS publications such as the house journal 'Population Trends', various 'Occasional Papers' and the series 'Studies on medical and Population Subjects'.

particular, the identification of those determinants which are susceptible to manipulation and which thereby may contribute to the prevention of disease.

The paragraphs which follow describe some such epidemiological studies, even if not explicitly undertaken by 'epidemiologists' the authors have used similar techniques for their given ends. In each example the objectives are the identification and description of the 'causes' of disease (or, more accurately, the risk factors associated with the disease), and the recommendation of policy to ameliorate these conditions.

There have been three major national perinatal surveys since the war. The first of these examined some 80,000 stillbirths and infant deaths among 1.5 million children born in England and Wales during 1949 and 1950 (Morris and Heady, 1955). Taking place immediately after the formation of the National Health Service in the post war years this study is significant as a basis for comparison with later surveys.

Findings from the 1949 study indicated that both 'social' (ie. social class, community, area of residence and so forth) and 'biological' (ie. maternal age, parity etc.) factors were associated with infant mortality and, importantly, that the two groups were highly interconnected. Thus,

"In the last resort social factors must act through the biological, ... while the importance of the biological factors for preventive medicine lies in the degree to which they can, if indeed they should, be controlled by individual and community action"

(Morris and Heady, 1955, p.348).

Far from showing an improvement in the rate of decrease of perinatal mortalities, the decade following this first survey saw a drop of only 3 per 1000. This compares with an average 1.5 per 1000 decrease in the perinatal mortality rate per year during and immediately after the war. In the light of this, the 1958

Perinatal Mortality Survey was instigated.

Using a 'case-control' approach, this had as its sample population 16,994 singleton births which took place in one week in March of that year, and 7,117 singleton stillbirths and neonatal deaths which occurred between March and May. The results of this study are reported in some detail in two volumes under the editorship of Butler et al. (Butler and Bonham, 1963; Butler, Alberman and Peel, 1969). The first report is of a general, descriptive nature, the objective being the rapid dissemination of the survey's findings rather than detailed analysis of the results. The effects of regional location, age, parity, social class, past obstetric history, place of booking and delivery, prenatal care, toxæmia, bleeding in pregnancy, gestational period, birthweight, labour and delivery (including breech, forceps and caesarian section) and pathology were all investigated in relation to perinatal mortality both in general and by specific causes. Relationships between most of these factors and the pregnancy outcome were found.

The second report, entitled 'Perinatal Problems', considers some of the more detailed aspects of the survey, for example the adverse effects of smoking during pregnancy, certain maternal factors such as primiparity, severe pre-eclampsia and short stature, and high risk predictors at booking and in pregnancy (all by the editors - Butler et al., 1969). Baird and Thomson (1969) looked at geographical differences in perinatal mortality as well as the effects of 'obstetric' and 'environmental' factors. Both of these were classified by the authors' own divisions of clinico-pathological cause. Gruenwald (1969) suggested an alternative classification of 'fetal' and 'extrafetal' factors affecting perinatal mortality.

The identification of a regional gradation in mortality experience with the highest rates in the north-west and the lowest rates in the south-east is of particular geographic significance. Baird and Thomson (1969) noted that these regional differences were paralleled by differences in the health, physique, reproductive habits and environment of the mother concerned and not necessarily related to standards of care. However, availability of care became important when considering the rural to urban differential. - distance from a specialist hospital centre makes it generally more difficult for rural dwellers to receive specialist antenatal care or to get to hospital should an emergency arise. These factors are, however, in part compensated for by "the superior health and physique of women in rural areas". (Baird and Thomson, 1969, p.253).

Only in the case of Aberdeen was an example given to show that a well-organised and efficient maternity service could reduce the perinatal mortality rate - chiefly by reducing 'avoidable' deaths to a minimum. In Aberdeen, nearly all women are confined in one teaching hospital under the supervision of specialist obstetricians and paediatricians, probably as a result of this the city has a favourable perinatal mortality rate.

The most recent national perinatal mortality survey was carried out in 1970, again under the joint auspices of the National Birthday Trust Fund and the Royal College of Obstetricians and Gynaecologists. This survey took place between 5th and 11th April 1970 and included all babies born, whether alive or dead, after the 24th week of gestation (ie. 4 weeks earlier than the 1958 survey), only deaths within the first week of life were subsequently analysed (the 1958 survey included deaths within one month). Altogether, 16,792 singleton births were included in the survey, of which 360 resulted in a stillbirth or

first week death, giving a perinatal mortality rate of just under 22 per thousand. (Claireaux,1975).

The results of the 1970 survey have been published in two volumes (R. Chamberlain et al., 1975 and G. Chamberlain et al., 1978). The first of these looks at the first week of life of the baby, while the second is more concerned with the obstetric care of the mother. The members of the Committee of the National Birthday Trust (1978) highlighted the following main conclusions.

First, the social class differential in mortality experience was still present and, if anything, had increased. Babies of unsupported mothers were found to be particularly at risk. Second, there had been little change in the proportion of low birthweight babies since the 1958 survey. This had important implications for obstetric practice since the chances of survival decrease sharply with declining birthweight and the likelihood of handicap increases. Third, although the perinatal mortality rate had fallen since 1958 for all mothers' age groups, the reduction in the more vulnerable groups (ie. mothers aged under 20 and over 35) had been less than that for the lower risk category. This in part explained the lack of change in the proportion of low birthweight babies - mothers aged less than 20 have been shown to give birth to a much greater number of small babies. Finally, concerning the risk to the unborn child from mothers who smoke cigarettes, not only were a higher proportion of mothers still smoking immediately after the birth of their baby but those that smoked more were least likely to give up. The second report of the 1970 survey found that

"women who smoked more than 5 cigarettes a day and developed pre-eclamptic toxæmia were at least 3 times as likely to lose their babies as non-smokers suffering the same complaint". (Chamberlain et al. 1978, p.262).

The implications of these results for medical practice were then considered by the committee. They observed that

"The numbers of perinatal deaths could be reduced substantially if maternity resources and medical expertise were better tailored to the needs of those mothers at greatest risk".
(Chamberlain et al., 1978, p.262).

The importance of high risk babies being born under specialist supervision and with immediate access to intensive care facilities was stressed.

Much important work has also been done at the local and regional scales. Examples include the ongoing Leicestershire survey (also described by MacVicar (1980)), an enquiry into perinatal deaths in the Mersey region (Mersey Region Working Party, 1982), the 1977 Scottish Perinatal Mortality Survey (McIlwaine,1980), and studies in Belfast (Elwood et al.,1974) and Cardiff (Chalmers et al.,1978). These studies varied with respect to sample size, population and survey method but many of their findings were in agreement both with each other and with the national surveys.

Among the factors most commonly found to be associated with perinatal mortality are included in the following:

- (i) Social class (Lambert,1976; Peckham et al., 1982; Adelstein et al., 1980; Macfarlane, 1979a)
- (ii) Legitimacy (Lambert,1976)
- (iii) Maternal age (Morris and Heady, 1955; Medical Statistics Division,1978; Butler and Bonham, 1963)
- (iv) Parity (Morris and Heady, 1955; Medical Statistics Division, 1978; Peckham et al., 1982)
- (v) Birthweight (Lambert, 1976;Baird, 1976;Elwood et al., 1974)
- (vi) Previous adverse outcome (Morris and Heady,1955; Butler and Bonham,1963; Elwood et al., 1974)

A wide range of other factors have also been suggested, these are summarised in Table 3.3. Some are of particular interest to the geographer, especially the socio-economic and spatial characteristics. Attention has already been drawn to the

Table 3.3 Factors associated with adverse perinatal outcome - a summary from the literature.

MATERNAL	INFANT
<p>(i) Physical Characteristics</p> <ul style="list-style-type: none"> * Age * Height or physique * Nutritional status * Year of birth * Alcohol consumption * Smoking habits <p>(ii) Past Obstetric history</p> <ul style="list-style-type: none"> * Parity * Birth interval * Abortion * Previous adverse outcome (stillbirth or neonatal death) * Previous complications (e.g. toxæmia, antepartum haemorrhage, caesarian section, premature birth) <p>(iii) Socio-economic characteristics</p> <ul style="list-style-type: none"> * Social class * Income or 'affluence' * Legitimacy * 'Social disadvantage' * Country of birth * Housing condition * Tenancy * Social 'stress' * Education * Employment during pregnancy * Family size * Paternal age <p>(iv) Spatial characteristics (applying to the mother's usual place of residence)</p> <ul style="list-style-type: none"> * Regional location * Urban or rural environment * Climatic conditions 	<p>(i) Characteristics of birth</p> <ul style="list-style-type: none"> * Length of labour * Gestational age * Type of delivery (e.g. breech, forceps etc.) <p>(ii) Other</p> <ul style="list-style-type: none"> * Birthweight * Multiplicity * Sex * Day of birth * Season of birth <p>HEALTH CARE</p> <p>(i) Antenatal care</p> <ul style="list-style-type: none"> * Availability * Accessibility * Quality * Early attendance <p>(ii) Delivery and postnatal</p> <ul style="list-style-type: none"> * Location of delivery (e.g. hospital, clinic or domiciliary) * Availability of specialist personnel and equipment

rural/urban differential in perinatal mortality experience in Britain, Nordstrom and Beckman (1981) describe a similar pattern in Sweden. They note that whilst birthweight normally increases after the first pregnancy, as exemplified by rural, non-polluted areas of northern Sweden, in nearby industrialised and urbanised areas this increase was much reduced. Among mothers occupationally exposed to air pollution at a smelter, birthweight in later pregnancies actually decreased.

The association of 'country of birth' or 'ethnic origin' with perinatal outcome is also of interest to the present study. Clarke and Clayton (1983) have already noted that in the Leicestershire survey to 1981 not only did Asian mothers have one and a half times the risk of perinatal mortality when social class, parity, height, legitimacy and their G.P.'s qualifications were taken into account, but also that Asian patients had a larger proportion of G.P.s that were not on the obstetric list (compared with non-Asians) which further added to the risk. Wharton et al., (1980), discussing the experience of the Asian mother at Sorrento Maternity Hospital in Birmingham, note that at birth Asian babies were on average smaller and lighter than the European equivalents. The authors suggest that these were probably not 'small normals' but that they had experienced pathological growth failure, due possibly to poor diet (for economic or religious reasons) during pregnancy. In Bradford, Lumb et al. (1981) found a persistently higher perinatal death rate for Asian babies born between 1974 and 1978 than for babies born to United Kingdom mothers. They noted that although some factors operate in favour of Asian women, such as fewer teenage mothers, lower illegitimacy and fewer smokers, many more presented an increased risk -

"...more women aged over 35, lower social class, higher parity, shorter pregnancy intervals, previous perinatal deaths, shorter duration of antenatal care, anaemia, shorter gestations, more

babies born without professional help, and more low-birthweight babies" (Lumb et al., 1981 p.106)

They suggest that

"encouraging expectant mothers to book early and attend regularly for antenatal care, to breast-feed their babies and space their pregnancies by at least one year, and to complete their families before the age of 40 should, if effective, help to diminish the PNMR for Asian patients in particular". (p.109)

The effect of ethnicity and of other variables on perinatal mortality in Leicester will be considered in Chapter 7. With respect to the geography of the perinatal problem, particular attention will be paid not only to the identification of individual risk characteristics, but also to the delimitation of areas of risk. It is therefore necessary first to take a broader view of the geography of Leicester.

Chapter 4. The Data Sources

4.1 Introduction

A major objective of this thesis is to show how data from initially incompatible sources can be combined to offer new insight into a medical geographical problem. Three main data sets are examined here : the 1981 Population census, the ongoing Leicestershire Perinatal Mortality Survey, and the 1980 Leicestershire Births Records. The perinatal survey offers detailed information on the social and health characteristics of the mother and infant, whilst the other data sets provide important background material describing the urban environment and the infant population in general. In view of the importance of all three data sets to this thesis this chapter describes each in detail. The limits of the study area are then defined.

4.2 The 1981 Census of Population

4.2.1 Introduction

The United Nations defines a census as

"the total process of collection, compiling, evaluating, analysing and publishing demographic, economic and social data pertaining, at a specified time, to all persons in a country or in a well-delimited part of a country"

(United Nations (1967), quoted by Dewdney (1981))

In Britain a general population census is taken every ten years, most recently in 1981. For a full account of this census the reader is referred to the comprehensive series of Working Papers produced by the Census Research Unit at the University of Durham, the Office of Population Censuses and Surveys (OPCS) Occasional Papers, and to the excellent "Census User's Handbook" edited by Rhind (1983). A summary of the points most relevant to the present research is given below.

Arguably the most important constraint to census taking today is cost. Rhind (1983, p.xvii) estimates that the 1981

census cost about £45 million and involved the employment of some 129,000 people. OPCS are therefore under tremendous pressure to maximise the usefulness of census data, an outcome which depends on a whole host of factors including the questionnaire design, the enumeration method, the sampling procedure and the efficiency of coding and validation. Of particular importance are the following considerations.

First, there is a critical balance between privacy and information (Norris, 1980). In Britain this is achieved through the dual mechanisms of question design and data aggregation. All census data are released on an areal basis, the smallest areal unit being the enumeration district (ED). Normally an ED will contain an average of 450 people, where fewer than 25 people (or 8 households) are present the data for that ED are suppressed, that is, only the total numbers of males and females are published. An advantage of suppression is that it eliminates the very small populations which would corrupt ratio type data values (Rhind and Dewdney, 1975). It does not affect the 10% sample data for socio-economic characteristics. Another procedure designed to protect the individual from disclosure of personal information is that of adjustment. This entails adding a 'small random error' to each individual count. The amount added may be -1, 0 or +1 (randomly generated in the ratio 1:2:1) but for each adjustment to an ED total, an inverse adjustment is made to a 'paired' ED total, with the result that the original total figures remain the same (Newman, 1978). Clearly the effect of this will be of greatest significance to the smaller EDs, for example those found at the edge of urban areas.

Secondly, there must be public acceptability of the questionnaire - a successful census depends on public cooperation. When a pilot survey was undertaken prior to the 1981

census, the inclusion of questions on ethnic origin caused such hostility that it was felt necessary to abandon all such questions, despite the vital need for such information. (Sillitoe (1978a and 1978b) describes this experiment). For similar reasons a set of questions pertaining to the fertility of married women were omitted from the 1981 census following a poor response in 1971. Both of these topics would have been relevant to this study of perinatal mortality.

Although continuity between censuses is important, so too is the need to reduce the amount of information recorded. Thus the 1981 census no longer required details of water supplies and kitchen facilities, the presence of a bath and flush toilet being considered satisfactory indicators of today's living standards. Again, the omission of information on matters potentially related to health and hygiene is unfortunate in the context of the present work.

The publication of the census results entails a complicated process of coding, validation and distribution. In Britain responses are manually transcribed from census forms to computer; checked for unacceptable or missing values; independantly validated through a combination of repeat enumeration, comparison with other data sources, and checks for internal consistency; and eventually disseminated in whichever format is required by the end user. For the first time, in 1981, data have been made widely available (at a cost) on computer tape along with suitable computer software with which to analyse it. The use of the data and the SASPAC package are described in Chapter 5.

4.2.2 The areal basis of the census

Although both postcodes and grid squares have been suggested for the compilation of census data (see Dewdney and Rhind, 1975;

Denham, 1980; Evans 1980, and Denham and Rhind, 1983 for a discussion), the fundamental spatial unit for data collection in England is the ED. Larger areal units, specifically electoral wards and local government districts, are built from aggregations of EDs. The definition of ED boundaries was a major task in planning the census, and was undertaken by both central and local agencies collaboratively. The following are some of the criteria (sometimes conflicting) which were applied: areas to be internally homogenous; boundaries to follow topographic features such as roads and railways but not to cross administrative boundaries such as wards, civil parishes or health districts; size to be determined by the workload for the enumerator; and lastly, inter-censal stability in the delimitation of boundaries. The final criterion was particularly difficult to meet, especially in urban areas. Changes in population distribution due to demolition and new building together with statutory boundary changes meant that as many as 56% of all EDs experienced boundary changes between 1971 and 1981.

4.2.3 Using census data

The most significant advantages and problems associated with the use of census data are summarised in Table 4.1. Although numerically there appear to be more problems than advantages, the relative importance attached to each will vary considerably between users. For the present study several of the problems are irrelevant (such as the timing of the census, and changes in questions and areal boundaries), while those that are significant are more than compensated for by the advantages of availability, coverage and cost. Of course certain problems are not even unique to the census - the ecological fallacy and modifiable areal unit problem are classic examples (see Chapter 5).

For an indication of the importance of the census to social

Table 4.1 Using Census Data

ADVANTAGES	PROBLEMS
Size of the data set	Timing - fixed
- national uniformity at small area level	- delay between enumeration and publication
- wide range of topics	Areal divisions - fixed for the purpose of census administration
High response rate	- unstable boundaries
Thoroughly planned, documented and supported	- irregular in size and shape
Wide availability	Definition of terms
High quality data	- some change between censuses
Continuity of some statistics	- complex census terminology
Cheap to users	Questions
'Objective neutrality'	- no control by individual research
(Hakim,1977)	- no information of a 'sensitive' nature
	- some lack of continuity between censuses
	Publication
	- available only in aggregate format
	- corruption of numbers due to confidentiality constraints
	Ecological fallacy and modifiable areal unit problems (notably the effects of changing scale from ED to ward level)
	Skewed distribution of many variables
	Combining census data with other data (spatially and temporally)

research, one needs to look no further than the vast quantities of census-based publications originating from central and local government, the health services, academia and commerce.

4.3 The Leicestershire Perinatal Mortality Survey

4.3.1 Context of the survey

The ongoing Leicestershire Perinatal Mortality Survey was instigated in 1975 following concern over an exceptionally high perinatal mortality rate apparent to mothers resident within Leicester city. As shown in Table 4.5 and Figure 4.1, between 1963 and 1975 the perinatal mortality rate for the city was almost always above both the national rate (England and Wales only) and the rate for the large urban areas. The Leicestershire county rate also continually exceeded the comparable rate for the rural districts. Despite an overall decline in each of these rates, that for Leicester city in 1975 showed no sign of relative improvement.

As explained in Chapter 3, the term 'perinatal mortality' refers to both a stillbirth and an early neonatal death. At the start of the survey it was observed that since the stillbirth rate for Leicester city did not deviate significantly from that for the other large urban areas (see Table 4.3), then the high perinatal mortality rate reflected an abnormally high early neonatal mortality rate (see Table 4.4) (MacVicar et al., 1976). Since no explanation could be offered for this, the need for a detailed review of early perinatal death was recognised.

The study had three main objectives:

- "(i) to identify in detail the factors contributing to the perinatal mortality rate in Leicestershire,
- (ii) to quantify any risk factors that were identified, and
- (iii) to assess the many factors which preceded each death, and to decide whether an alternative action by any individual,

Table 4.2 Perinatal Mortality: National and Regional Comparisons

Total Live Births

	England+ Wales	Urban Areas(1)(2)	Rural Districts(3)	Leicester- shire(4)	Leicester C.B.(5)
1960	785005	100850	155114	11644	4540
1961	811281	103868	159788	12291	4721
1962	838736	112204	166908	12900	5057
1963	854055	113963	172064	13023	5016
1964	875972	114159	178559	13356	5098
1965	862725	108268	178757	13314	5002
1966	849823	106226	176995	13268	5146
1967	832164	105437	173769	13220	5267
1968	819272	110215	171418	13278	5143
1969	797538	106907	170147	13192	5118
1970	784486	104189	168274	13079	4862
1971	783155	103792	170756	13185	4756
1972	725440	94792	161192	12237	4385
1973	675953	87612	152047	11597	3958

	England+ Wales	Trent R.H.A. (6)	Leicestershire A.H.A.	Leicester C.D.
1974	639868	59033	11271	3850
1975	603448	55623	10761	3682
1976	584270	53800	10610	3769
1977	569323	52293	10639	3841
1978	596461	54377	10846	3875
1979	638028	58018	11563	4247
1980	656234	60356	12194	4664
1981	634456	58111	11648	4527

- (1) Includes only urban areas outside conurbations with populations of 100,000 and over at the 1961 census
- (2) Boundary changes in 1968
- (3) Boundary changes in 1968
- (4) Boundary changes in 1965 and 1966
- (5) Boundary changes in 1966
- (6) From 1978 includes part of North Derbyshire

Table 4.3 Perinatal Mortality : National and Regional Comparisons

Stillbirths

	England+ Wales	Urban Areas ^{(1) (2)}	Rural Districts ⁽³⁾	Leicester- shire ⁽⁴⁾	Leicester C.B. ⁽⁵⁾
1960	15819 (19.8)	2067 (20.1)	3031 (19.2)	231 (19.5)	105 (22.6)
1961	15727 (19.0)	2068 (19.5)	3016 (18.5)	228 (18.2)	76 (15.8)
1962	15464 (18.1)	2110 (18.5)	2999 (17.7)	229 (17.4)	92 (17.9)
1963	14989 (17.2)	2076 (17.9)	2863 (16.4)	241 (18.2)	99 (19.4)
1964	14546 (16.3)	1999 (17.2)	2784 (15.4)	227 (16.7)	96 (18.5)
1965	13841 (15.8)	1785 (16.2)	2652 (14.6)	203 (15.0)	69 (13.6)
1966	13243 (15.3)	1727 (16.0)	2548 (14.2)	213 (15.8)	85 (16.2)
1967	12528 (14.8)	1678 (15.7)	2441 (13.9)	205 (15.3)	89 (16.6)
1968	11848 (14.2)	1677 (15.0)	2300 (13.2)	195 (14.5)	93 (17.8)
1969	10654 (13.2)	1487 (13.7)	2005 (11.6)	178 (13.3)	74 (14.3)
1970	10345 (13.0)	1464 (13.8)	2100 (12.3)	146 (11.0)	61 (12.4)
1971	9899 (12.5)	1386 (13.2)	2004 (11.6)	152 (11.4)	62 (12.9)
1972	8799 (12.0)	1238 (12.9)	1752 (10.8)	137 (11.1)	62 (13.9)
1973	7936 (11.6)	1129 (12.7)	1600 (10.4)	143 (12.2)	60 (14.9)

	England+ Wales	Trent R.H.A. ⁽⁶⁾	Leicestershire A.H.A.	Leicester C.D.
1974	7175 (11.1)	721 (11.9)	114 (12.6)	54 (13.8)
1975	6292 (10.3)	567 (10.1)	124 (11.4)	44 (11.8)
1976	5709 (9.7)	561 (10.0)	132 (12.3)	57 (14.9)
1977	5400 (9.4)	472 (8.9)	94 (8.8)	45 (11.6)
1978	5112 (8.5)	446 (8.1)	107 (9.8)	40 (10.2)
1979	5125 (8.0)	469 (8.0)	97 (8.3)	50 (11.6)
1980	4773 (7.2)	427 (7.0)	79 (6.4)	34 (7.2)
1981	4203 (6.6)	360 (6.2)	54 (4.6)	25 (5.5)

Sources:

(a) 'The Registrar General's Statistical Review of England and Wales', for the years 1960 to 1973. Part 1. Medical tables. H.M.S.O., London.(Until 1968 produced by G.R.O.then by O.P.C.S.).Table 13

(b) Leicestershire A.H.A. Statistics and Information Unit. 'Information Profile: Vital Statistics for Leicestershire'. (Various years).

Table 4.4 Perinatal Mortality : National and Regional Comparisons

Deaths under 1 week (Early neonatal deaths)

	England+ Wales	Urban Areas ⁽¹⁾⁽²⁾	Rural Districts ⁽³⁾	Leicester- shire ⁽⁴⁾	Leicester C.B. ⁽⁵⁾
1960	10475 (13.3)	1390 (13.8)	1919 (12.4)	149 (12.8)	70 (15.4)
1961	10768 (13.3)	1481 (14.3)	1895 (11.9)	146 (11.9)	69 (14.6)
1962	10888 (13.0)	1497 (13.3)	2053 (12.3)	135 (10.5)	57 (11.3)
1963	10498 (12.3)	1463 (12.8)	2016 (11.7)	159 (12.2)	67 (13.4)
1964	10537 (12.0)	1354 (11.9)	1975 (11.1)	137 (10.3)	70 (13.7)
1965	9732 (11.3)	1293 (11.9)	1760 (9.8)	150 (11.3)	61 (12.2)
1966	9446 (11.1)	1230 (11.6)	1801 (10.2)	156 (11.8)	74 (14.4)
1967	8945 (10.7)	1172 (11.1)	1658 (9.5)	152 (11.5)	73 (13.9)
1968	8682 (10.6)	1170 (10.6)	1627 (9.5)	111 (8.4)	53 (10.3)
1969	8229 (10.3)	1121 (10.5)	1581 (9.3)	127 (9.6)	61 (11.9)
1970	8326 (10.6)	1083 (10.4)	1608 (9.6)	128 (9.8)	50 (10.3)
1971	7750 (9.9)	1061 (10.2)	1618 (9.5)	135 (10.2)	61 (12.8)
1972	7144 (9.8)	985 (10.4)	1430 (8.9)	137 (11.2)	64 (14.6)
1973	6438 (9.5)	865 (9.9)	1293 (8.5)	129 (11.1)	53 (13.4)

	England+ Wales	Trent R.H.A. ⁽⁶⁾	Leicestershire A.H.A.	Leicester C.D.
1974	5987 (9.4)	536 (9.1)	112 (9.9)	45 (11.7)
1975	5462 (9.1)	474 (8.6)	85 (7.9)	40 (10.9)
1976	4756 (8.1)	485 (9.0)	98 (9.2)	45 (11.9)
1977	4353 (7.6)	407 (7.8)	94 (8.8)	35 (9.1)
1978	4241 (7.1)	392 (7.2)	83 (7.7)	33 (8.5)
1979	4307 (6.8)	389 (6.7)	78 (6.7)	38 (8.9)
1980	4034 (6.1)	363 (6.0)	73 (6.0)	37 (7.9)
1981	3354 (5.3)	299 (5.1)	66 (5.7)	33 (7.3)

Notes: Calculation of rates (figures in parentheses)

Stillbirth rate = no. of stillbirths per 1000 live and stillbirths

Early neonatal death rate = no. of early neonatal deaths per 1000 live births

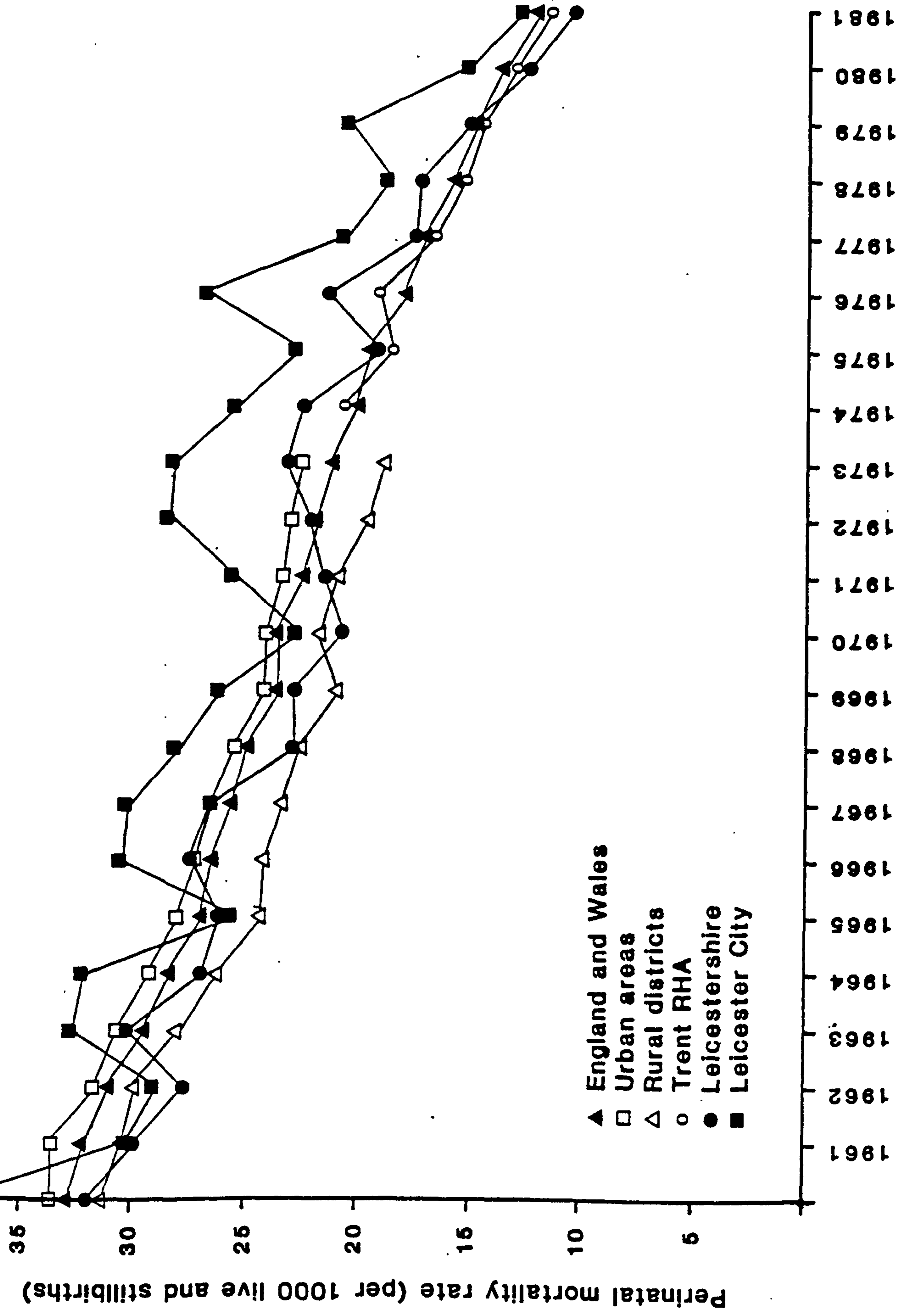
Perinatal mortality rate = no. of stillbirths plus no. of early neo-natal deaths per 1000 live and stillbirths.

Table 4.5 Perinatal Mortality : National and Regional Comparisons

Perinatal Mortality Rate

	England+	Urban Areas ^{(1) (2)}	Rural Districts ⁽³⁾	Leicester- shire ⁽⁴⁾	Leicester C.B.
~~~~~					
1960	32.8	33.6	31.3	32.0	37.6
1961	32.0	33.5	30.1	29.9	30.2
1962	30.8	31.6	29.7	27.7	28.9
1963	29.3	30.5	27.9	30.2	32.5
1964	28.1	28.9	26.2	26.8	32.0
1965	26.8	28.0	24.3	26.1	25.6
1966	26.3	27.4	24.2	27.4	30.4
1967	25.4	26.6	23.3	26.6	30.2
1968	24.7	25.4	22.6	22.7	27.9
1969	23.4	24.1	20.8	22.8	26.0
1970	23.5	24.1	21.7	20.7	22.5
1971	22.3	23.3	20.9	21.5	25.5
1972	21.7	23.1	19.5	22.1	28.3
1973	21.0	22.5	18.8	23.2	28.1
~~~~~					
	England+ Wales	Trent R.H.A. ⁽⁶⁾	Leicestershire A.H.A.	Leicester C.D.	
~~~~~					
1974	20.3	20.7	22.4	25.4	
1975	19.3	18.6	19.2	22.5	
1976	17.7	19.2	21.4	26.7	
1977	17.0	16.7	17.5	20.6	
1978	15.5	15.3	17.3	18.6	
1979	14.7	14.7	15.0	20.5	
1980	13.3	13.0	12.4	15.1	
1981	11.8	11.3	10.3	12.7	
~~~~~					


Figure 4.1 National and regional perinatal mortality rates



including the patient, would have reduced the likelihood of the death." (MacVicar et al., 1977, p.6).

4.3.2 Choice of method and survey population

The retrospective case/control method of investigation (see Chapter 3) was chosen on the basis of cost and practical considerations. Two problems arose. The first of these, that of extracting information from medical records, was overcome by the use of the specially designed questionnaire (see Section 4.3.3). The second, the selection of an appropriate control population, is considered below. The 'case' group comprised all infants born to Leicestershire mothers who failed to survive the perinatal period.

The control group ideally included a representative sample of all live births. This may have been taken retrospectively, say from routine births records (see Section 4.3.4 for an example of such a data source), or defined concurrently with the case population. Since the former method involved waiting until these data were completed (for births records this occurs only at the year end) the resulting delay potentially caused various problems, for example changes of address, differential recall or subsequent pregnancy. As a result the latter method was preferred. A systematic selection of every 'nth' birth was considered but rejected in favour of a paired case/control system based on a geographical criterion. Thus, following a perinatal death, the control selected was the next live birth occurring at the intended place of delivery of the case (Clarke and Clayton, 1981). Use of the 'intended' rather than actual place of delivery ensured that all control mothers could be interviewed in Leicestershire hospitals.

4.3.3 The Questionnaire

The questionnaire was designed in line with the study's main

objectives (see Section 4.3.1), although it has been modified several times in the course of the survey, to better reflect the generation and testing of secondary hypotheses (Clarke and Clayton, 1981, see Appendix 3.2), by 1983 in outline it was as follows.

The first of the four sections was based on the mother's case notes and cooperation card; it included questions on the mother's age, status, past and present medical and obstetric history, her antenatal care and labour, and the infant's condition at birth.

Section two was completed at an interview with the mother when her condition permitted. She gave details of her own and the father's occupation and social background, the drugs she took during the pregnancy, and the frequency with which she consumed various types of food, alcohol and cigarettes.

Part three was concerned with the cause of the infant's death. Both direct and antecedent causes, as well as 'other significant conditions' were included. This section could be modified following post mortem and discussion of the case.

The final part was concerned with whether the death had been avoidable, this being assessed by a panel of medical staff involved in the case. An 'avoidable' death was defined as either one in which the quality of medical care during the pregnancy fell below the generally accepted standard, or one in which the patient failed to give full cooperation.

The questionnaire is reproduced in Appendix 3.1.

4.3.4 Published findings

In keeping with the objectives of the study, the annual reports produced by Leicester University's Department of Community Health have focussed on the changing overall rates of perinatal mortality in the county; the variations in perinatal experience

among the Asian population subgroups (Sikhs, Muslims and Hindus); and the effects of changing medical provision (examples including the availability of the A.F.P. (alpha-fetoprotein) screening programme, and the establishment of new antenatal clinics in high risk areas). The dissemination of experiences and results to other health regions has been a priority.

The emphasis in this work has been primarily on the medical and health care aspects of the study. Although the spatial pattern of mortalities has been considered (for example by Burton (1982)), there is still an expressed need to consider this aspect in more detail. In particular there is a need to consider the perinatal problem within its wider social and economic context.

4.3.5 Using survey results in a geographical context

Although the value of epidemiological data such as the Perinatal Mortality Survey should not be underestimated, there are several drawbacks to their use in geographical work. First, the data have been collected by medical personnel for medical rather than geographical purposes. Social, economic and spatial references are included in the perinatal survey questionnaire but the range of questions is limited and does not include information on, for example, tenancy, quality of housing, availability of transport and so forth.

Secondly, the necessity for confidentiality restricts the use of epidemiological data. This applies both to the type of information collected and to the manner in which it is made available. In the present study the postcode of the patient's home address was offered as the only spatial reference. Since a postcode may refer to up to 70 addresses and normally covers part of a postman's daily route, the representation of the resulting elongated area by a single point must be viewed as a

generalisation of the mother's actual place of residence. This applies whether the point is a centroid or 'start of walk' position.

Thirdly, the usefulness of the data is dependant on decisions taken in the preliminary stages of data collection and on the recognition of any constraints which may result. For example, in the present survey the effect of hospital referral policy with regard to pregnant women is unrecognised even though this determines the intended place of delivery and hence the selection of the control population. Thus, if all pregnant mothers were referred to their nearest maternity hospital, within the resulting 'catchment areas' numbers of controls would by definition exactly equal numbers of cases. That this does not occur is probably due to a separate referral policy for 'high risk' births. MacVicar et al. (1976) have demonstrated that some peripheral units refer complicated pregnancies to one of the two major city hospitals (the Leicester Royal Infirmary Maternity Hospital and the Leicester General Hospital), while Clarke and Clayton (1981) observe that, because of these policies, these hospitals have high proportions of zero and high parity mothers, and those in the older age groups. As may be expected, the presence of a disproportionate number of 'high risk' patients results in both a higher perinatal mortality rate and, significantly, a high risk 'pool' of mothers from which to draw the control population. The net effect of the referral policy is therefore to weaken the validity of any case/control comparison.

4.4 The Leicestershire births records

Since 1973 the Leicestershire Area Health Authority, as part of their vaccination and immunisation programme, have kept individual computerised records for each child born in the county. Upon each birth a short questionnaire is completed from which the

following details are coded: a code number for the birth, birth outcome, birthweight, maternal age, ethnic group, the screening clinic attended and the postcode of the mother's home address. Other information pertaining to the mother's medical and obstetric history, her health care during pregnancy, the condition of the infant at birth and each parent's occupation are included on the questionnaire but not coded. The purpose of using this ancillary data source is to 'standardise' the perinatal mortality data. Standardisation in medical geography usually involves converting an absolute value for some morbidity or mortality indicator into a rate per population 'at risk' (see Chapter 2). In the case of a perinatal death the 'at risk' population comprises all newly born infants. The births data therefore perform two functions: to standardise the mortality data, in particular with regard to its spatial distribution, and to assess the representativeness of the control population.

Due to the large number of births involved (some 11,000 annually) only those for one year were processed for this study. The year chosen was 1980, to correspond with the imminent 1981 census. A total of 12,190 births were recorded, of which 6,302 were to mothers residing in the Greater Leicester study area.

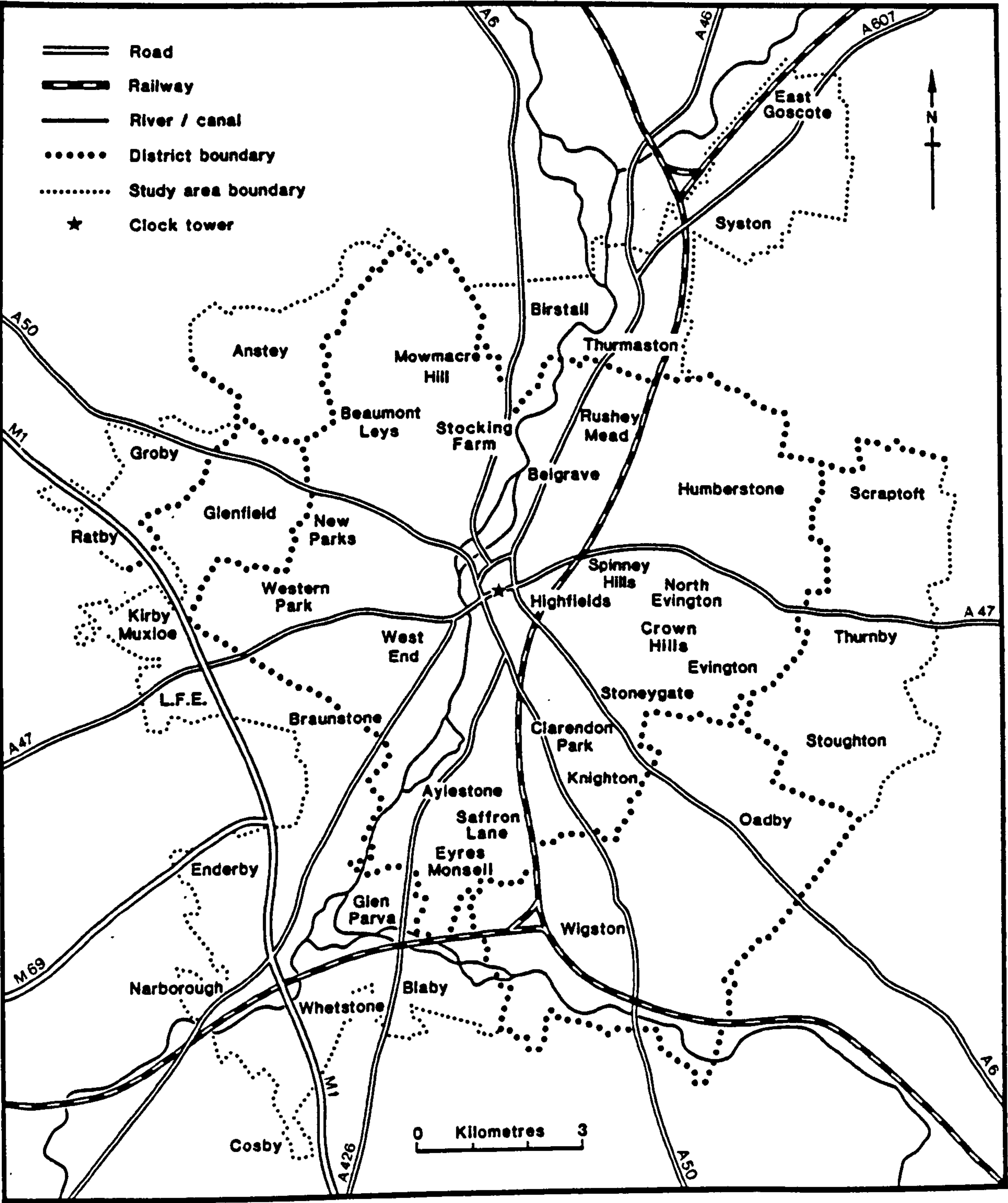
4.5 Choice of study area

Following the ongoing trend towards suburbanisation and the increasing proportion of people living in the outer areas of Leicester it was decided that the study area should encompass not only the city and administrative area, but also the contiguous built-up area surrounding this. The exact limits were defined by ED boundaries to facilitate analysis using census data.

Figure 4.2 shows the extent of Greater Leicester with the main settlements marked. As it indicates, the area covered is some

Figure 4.2

Greater Leicester – settlements and communications



166 square kilometres, extending from Birstall, Syston and Goscote in the north, to Oadby and Wigston in the south. The most westerly village is Kirby Muxloe while to the east the area extends as far as Scraptoft and Thurnby. In total, some 847 EDs were included.

Chapter 5. Classifying Census Areas

5.1 Introduction

This chapter outlines an attempt to classify EDs on the basis of attributes from the 1981 census. The objectives underlying the classification were four-fold:

- to identify classes of EDs with similar social, economic and demographic characteristics about which general statements may be made
- to define 'social areas' based on aggregations of contiguous EDs which may be used as single units for subsequent analysis
- to generate hypotheses regarding the differences between these areas and, in particular, to allow unexpected and unpredicted associations to be revealed
- to enable preliminary testing of such hypotheses.

Two other practical benefits were expected to accrue. The first of these was relevant to the distribution of the population and the nature of the census data. Namely, that by aggregating EDs on the outskirts of the city problems of small sample size could be expected to be avoided. The second was the advantage of having fewer areas for subsequent processing and analysis.

A definition of purpose is an essential pre-requisite of any classification. In the first instance this will determine the choice of both method and variables. (These are discussed in detail later in this chapter). The variables chosen were transformed into component scores following a principal components analysis. The method chosen was a cluster analysis using an agglomerative procedure. In this the practice is normally to

"start with a lot of individuals and proceed, using a set of pre-determined rules, to allocate them to groups according to levels of similarity on the chosen criteria" (Johnston, 1976b, p.4).

Having created a first set of groups, these may themselves be combined on the same principles. A logical end is reached when all individuals are part of a single group. Both hierarchic (entailing irrevocable fusion of cluster members) and non-hierarchic

(allowing subsequent reallocation) methods were tried.

5.2 Problems of classifying areal data.

Although undoubtedly an important and useful stage in scientific investigation, the process of classification is not without criticism which arises mainly from the problems inherent in the technique - see for example Craig (1975) and Openshaw et al. (1980). Problems of classifying areal data in general, and census data in particular, fall into four basic categories, those associated with the data themselves, the methodology and statistical procedures used, the operational decisions taken, and the interpretation of the results.

At a general level, data problems include volume, timing and within-area variation. Multivariate classifications frequently involve a large number of observations (variables) for an even larger number of areal units. Openshaw et al's (1984) 60 variable classifications of 78,041 EDs and 15.627 million households from the 1971 Italian census are extreme examples, but Webber's (1979) classification is typical, involving 40 variables and 3,996 EDs.

It is self-evident that a classification is applicable only to the region from which the original data were collected, it is also true that it applies to that region only at the time of the enumeration. This is of particular concern in areas of rapid social change. The distribution of the population and the delimitation of the original area boundaries must be considered, especially if there is a risk of the EDs containing populations lacking common characteristics, in which case the 'quality' of the data and its suitability for subsequent aggregation may then be a problem. Nevertheless it is also important to bear in mind the criteria upon which groupings are based since wide within-area variations in crime rates for example, would not seriously affect the validity of a classification based on housing type.

Methodological and statistical problems. These frequently arise together. A common criticism is that a method or statistical technique is chosen because of the availability of a convenient computer program and not for carefully thought out reasons. The ease of use of many packages also encourages disregard of the underlying statistical principles. For example, any classification technique which relies on the calculation of distance matrices assumes that the variables involved are uncorrelated (or orthogonal) (Gatrell and Cole, 1983). If the classification uses principal component scores for its data set then this is not a problem, but raw, untransformed data would rarely fulfil this condition.

It has been argued (Openshaw et al., 1980) that statistical classification methods are by nature exploratory techniques whose results are dependant on the 'arbitrary' decisions taken during the process. Openshaw and Gillard (1978) have examined in detail some of the effects of changing certain operational decisions on a spatial classification. Their reservations are recognised here, but they were not felt to be helpful to the present study since the focus was not on the instability and lack of robustness (and hence unhelpfulness) of spatial classifications, but rather on the more positive aspects of their utility. In principle, it was felt that if each decision was taken with due consideration for the pre-defined objectives, then the resulting classification would be interpretable, and hence acceptable, in terms of those objectives.

There are two types of operational decision, those that are made in the preliminary stages of the analysis, and those which may be termed 'methodological decisions'. Preliminary decisions include the choice of variables - their number, type and, if applicable, their weighting. Depending on the method used the

number of classes may also have to be decided prior to the analysis, as might the size of each class. The number, type and size of individual areas is a third preliminary decision. At this stage, for example, it may be decided that EDs with a specified minimum population should be excluded from the analysis.

Decisions taken during the classification process are far more numerous and equally important to the final result, for example, type of data, type of classification, number of classes, the need for a contiguity constraint and so forth. These decisions affect both the result and its interpretation. Whilst objectivity here is desirable, it can only be achieved insofar as methods are capable of independent replication (Openshaw, 1982d). Over-interpretation of the results is a common danger - for example applying the conclusions of a classification to a different scale of area or to the individual within a classified area. To avoid this it is important to ask 'is this general classification meaningful and valid in the context of my data?' and 'what do these clusters mean to the people who live in them?'.

The 'ecological fallacy' and the 'modifiable areal unit problem' (MAUP) have received much attention in the literature over recent years (see Jones 1982, Openshaw 1982b or Openshaw 1982c). They are defined as follows: the ecological fallacy is that the results for aggregate data can be generalised to individuals. The MAUP arises from a combination of scale and aggregation problems, in turn the consequence of individuals being classified into either different numbers of areas or different arrangements of areas at the same scale.

Robinson (1950) first suggested that statistical associations for aggregated populations might differ in magnitude and even in sign from those for individual population members. More recently the full implications of this have become apparent

to geographers with recognition of the MAUP. In an examination of ecological fallacies in the analysis of census data, Openshaw poses the following questions:

"Do the results provided by Robinson (1950) apply to all variables or merely to some variables?"

"Do they apply to all aggregated data or only to the more grossly aggregated data?"

"Can the more sensitive variables be identified from their statistical and geographical distributions?"

"Are some analytical techniques more likely to amplify the differences between individual and aggregated level results?" (Openshaw, 1982b, p.3).

The same questions are pertinent to an evaluation of the MAUP.

Both the ecological fallacy and the MAUP are relevant to the present study. The classification of census areas and the identification of areas of risk from the perinatal data both depend on some level of aggregation and generalisation. It is suggested however, that the effect of both problems can be minimised, given the following conditions. First that the objective of the classification was to create meaningful new small areas, these areas then became the objects of interest and no attempt was then necessary to apply results to either the individual level or to other sets of areas, Second, both the ecological fallacy and the MAUP mainly arise at the interpretation stage and can be avoided by being aware of both problems and by resisting the temptation to 'over-interpret' the results.

In conclusion for this study, classification was viewed as a means to an end and not an end in itself. The advantage the technique offers in being able to summarise this complex multivariate data set more than compensates for the problems encountered.

5.3 Choosing variable for the analysis.

The importance of selecting appropriate variables prior to

any form of multivariate analysis has already been mentioned. This section explains some of the criteria upon which the choice of variables was made, considers some of the similar experiences of other projects and describes the main features of the data set used here.

Various authors have stressed the importance of a clear statement of purpose (Visvalingam,1981) as a first step in the identification of census indicators. In this case the objective of the analysis was to provide an accurate description of the Greater Leicester area (as defined in Section 4.5) on the basis of the demographic, social and economic characteristics of its population. This was to provide the basis for classifying EDs within the study area into larger homogenous units, units which were to be used as a foundation for more specific work on the health experiences of the resident population. Since the classification was seen as a hypothesis-generator and starting point for future work, a general rather than purposive classification, and hence set of variables, was appropriate.

In selecting variables the following considerations and principles were taken into account (see the work of Hakim (1978), Evans (1980) and Rhind et al. (1977) for a similar discussion). First, of the four types of social indicator identified by Carlisle (1972) - informative, predictive, problem-oriented and programme-evaluation - only informative (or descriptive) and problem-oriented (or diagnostic) indicators are relevant in the context of the present study. The former include such variables as 'percentage owner-occupied housing' and are essentially descriptive of social conditions, while the latter are exemplified by 'percentage of household without exclusive use of all amenities' which is indicative of housing stress.

Secondly, a balance needed to be achieved between the different variables with respect to their explanatory power in the analysis. Particular attention was paid to the level of correlation between variables. On the one hand, all but one of any set of highly correlated variables were, if possible, either combined or omitted to ensure that the analysis was not biased toward that set of factors. On the other, care was taken not to combine negatively correlated data cells. (Evans (1980) notes that the 15-44 year age group is inappropriate for this reason).

Two types of variable were felt to be of special interest. Those that were likely to discriminate effectively between areas, such as the proportion of council households, and those which, still within the context of a 'general' classification, were able to identify 'problem' areas and conditions. 'Overcrowding' and shared facilities were examples of the latter.

Potential problems arise when using census data because of the need to hold variables as ratio values, frequently within the range 0% to 100%. In addition to the statistical problems encountered in processing 'closed' data sets (see Evans (1981) for a full discussion), the choice and size of denominator are critical since ratios are very susceptible to changes in either of these. In the present work the most commonly used denominators were 'private households' and 'residents in private households'. Where several variables formed subsets of the same total, these were given the same denominator. The visiting or 'de facto' population was not included.

The demands of the more sophisticated statistical techniques usually include a statement on the statistical distribution of the variables. Unfortunately many important census indicators are known to have skewed distributions (Rhind, 1975). If these cause sufficient concern, any potential problems can usually be overcome

by using transformations (see Roff,1977).

With regard to previous census work, Hakim (1977) lists no less than 68 census-based area profiles produced mainly by local authorities, academics and central government. A representative selection of these are considered here. Table 5.1 summarises the number and type of variables each includes, a balance is important since it will determine the outcome of the analysis. This does not mean, however, that each class of indicator must be of identical size even for a 'general' classification, indeed, the availability of counts within the census itself would make this difficult to achieve.

'Demographic' variables refer to the age and sometimes the sex structure of the population. Either all or selected age groups may be included. For example Imber's (1977a) study uses only two age groups - young children (aged 0-4) and the elderly (aged 65 and over). Both of these groups place particular demands on health and social services and are thus indicators of need in addition to being demographic statistics. Webber (1977c) included the marriage rate as a demographic index.

Census users are most frequently interested in the extreme types of **household composition** such as households with two adults and five dependant children (Webber, 1975), or those which may be judged to have special needs, for example households composed of lone parent families with dependant children (Webber and Craig,1978). Additionally, household size, the proportion of pensioners and fertility rates are examples of household composition variables.

In each of the studies described, indicators of **housing status** form the largest and most diverse group. Housing tenure, overcrowding, amenities, housing space and dwelling size form a

Table 5.1 Census variables used in multivariate classifications

Author(s)	Area of Study	Number of Variables included in study								
		Demographic	Household Composition	Housing	Socio-economic structure	Employment	Ethnicity	Migration	Transport	Total
Imber(1977)	England	2	3	6	3	6	2	-	1	23
Webber and Craig (1978)	England and Wales	7	5	10	6	9	1	1	1	40
Webber (1979)	Great Britain	6	4	13	5	6	1	1	4	40
Daly (1971)	Greater London	1	-	4	2	-	2	1	1	11
Webber (1975)	Liverpool	6	4	12	6	6	1	1	4	40
Webber (1977a)	Cumbria	7	3	12	5	7	0	1	5	40
Webber (1977b)	Haringey	7	3	17	5	5	1	1	1	40

common core. Serving mainly as 'deprivation indicators', these variables must be carefully selected from the 1,330 counts for housing data in the 1981 census. In particular, appropriate 'cut-off' points should be designated. For example, thirty years ago a population density figure of 1.5 persons per room might have been acceptable, today it represent acute overcrowding (Hakim,1978). Similar awareness must be shown when deriving variables representing 'large' families, 'large' dwellings, the 'elderly', etc.

Indicators of **socio-economic structure** comprise mainly a subset of the conventional socio economic groups. Webber uses a five-category breakdown in most of his studies, namely, the proportion of heads of household in professional/managerial, non-manual, skilled manual, semi-skilled manual and unskilled occupational groups. Daly (1971) on the other hand, includes only employer/managers and skilled workers, both measured as a proportion of economically active males, whilst Imber (1977a) includes only the proportion of the total population that are unskilled workers. There is no consensus of opinion regarding either the sub-groups to be included, or the most useful denominators for the variables.

Other variables which may be classed as representative of socio-economic structure include, for example, educational attainment and the work rate among married women. The fact that the latter might also be included under 'household composition' or 'employment' serves to highlight the interdependant nature of this classification of variables.

Closely related to the previous category, **employment** variables normally refer to engagement in the different industries, (for example agriculture, manufacturing/mining and services used by Webber,1979), but they may also include rates of

unemployment, economic inactivity, sickness amongst males and female employment. The last of these is seen as being indicative of social change as increasing numbers of women go out to work. Hakim (1978) suggests that female unemployment tends to correlate more significantly with malaise indicators than that of males, and consequently that economic activity rates for women may be used as an indicator of local prosperity. Conversely, male unemployment may be used as an indicator of industrial decline or be indicative of a phase of transition in industrial character.

The 1981 census differs from earlier enumerations in that the only indicator of **ethnicity** is the household head's country of birth, which is tabulated by sex, age and birthplace (inside or outside the UK) of persons in the household. This statistic has the two disadvantages of being unable to identify either white children born overseas to British parents, or British born children of immigrants. Migrancy rates are similarly difficult to determine, with the best estimate coming from the proportion of the population changing address within the twelve months prior to census night. This can only serve as a surrogate measure since it cannot describe adequately the dynamics of population movement.

Means of travel to work and car ownership are the most common '**transport**' variables extracted.

Following consideration of all the above points, an initial set of 39 variables was selected (see Table 5.2). These comprised 7 demographic variables, 6 household composition, 8 housing, 6 socio-economic structure, 7 employment, 3 ethnicity, 1 migration and 1 transport variable.

5.4 A classification of enumeration districts in Greater Leicester - methodology.

Given the importance of operational decisions to the classification process, it was decided that a fully automated

Table 5.2 Classifying enumeration districts - census variables.

Variable	Name	Calculation*
1. DEMOGRAPHIC INDICATORS		
Young children	DYNGCH	$\frac{\text{All residents aged 0-4}}{\text{Total residents}}$
Older children	DOLDCH	$\frac{\text{All residents aged 5-15}}{\text{Total residents}}$
People of working age - young group	DYNGWK	$\frac{\text{All residents aged 16-24}}{\text{Total residents}}$
People of working age - middle group	DMIDWK	$\frac{\text{All residents aged 25-44}}{\text{Total residents}}$
People of working age - older group	DOLDWK	$\frac{\text{All residents aged 45-59} + \text{plus men aged 60-64}}{\text{Total residents}}$
People of pensionable age - young group	DYNPNS	$\frac{\text{All residents aged 65-74} + \text{plus women aged 60-64}}{\text{Total residents}}$
People of pensionable age - older group	DOLPNS	$\frac{\text{All residents aged 75+}}{\text{Total residents}}$
2. HOUSEHOLD COMPOSITION		
Single person households	HHSING	$\frac{\text{Private households with 1 person only}}{\text{Total private households with residents}}$
Large households	HHLRGE	$\frac{\text{Private households with more than 5 persons}}{\text{Total private households with residents}}$
Many dependant children	HHCHIL	$\frac{\text{Private households with 4 or more persons aged 0-15}}{\text{Total private households with residents}}$
Lone parent households	HHLPAR	$\frac{\text{Lone adults resident in private households with residents aged 0-15}}{\text{Total private households with residents aged 0-15}}$
Pensioner households	HHPENS	$\frac{\text{Private households containing only persons of pensionable age (any number)}}{\text{Total private households with residents}}$

***Note:** all calculations are multiplied by 100 to give the ratio as a percentage.

Table 5.2 (cont.)

Variable	Name	Calculation
Mixed families	HHXFAM	Private households with 3 or more adults (any combination) and 1 or more persons aged 0-15 <u>Total private households with residents.</u>
3. HOUSING INDICATORS		
Owner-occupied	HSOWN	<u>Owner-occupied households</u> Total private households with residents
Council property	HSCOUN	<u>Council etc. households</u> Total private households with residents
Private rented	HSRENT	'Other' rented unfurnished plus other rented furnished <u>Total private households with residents</u>
Lacking full amenities	HSAMEN	Private households without exclusive use of bath and inside w.c. <u>Total private households with residents</u>
Overcrowding	HSCRWD	Private households with more than 1.5 persons per room <u>Total private households with residents</u>
Bedsits	HSBEDS	<u>Household spaces in bedsits</u> Resident household spaces in permanent buildings
Shared dwellings	HSSHAR	Private households not in self-contained accommodation <u>Total private households with residents</u>
Flats	HSFLAT	<u>Household spaces in purpose-built flats</u> Resident household spaces in permanent buildings
4. SOCIO-ECONOMIC STRUCTURE (10% sample)		
Professional and managerial headed households	SEPROF	Households with heads of socio-economic groups 1,2,3,4 and 13 <u>Total private households with residents</u>
Non-manual headed households	SENMAN	Households with heads of socio-economic groups 5 and 6 <u>Total private households with residents</u>

Table 5.2 (cont.)

Variable	Name	Calculation
Skilled manual headed households	SESMAN	$\frac{\text{Households with heads of socio-economic groups 8,9,12 and 14}}{\text{Total private households with residents}}$
Semi-skilled manual headed households	SESS	$\frac{\text{Households with heads of socio-economic groups 7,10 and 15}}{\text{Total private households with residents}}$
Unskilled headed households	SEUS	$\frac{\text{Households with heads of socio-economic group 11}}{\text{Total private households with residents}}$
Highly educated	SEEDUC	$\frac{\text{Residents aged 18 or over with degrees, professional or vocational qualifications}}{\text{Total residents (10\% sample)}}$
5. EMPLOYMENT INDICATORS		
Economically active	EMEACT	$\frac{\text{All economically active persons aged 16 or more}}{\text{Total residents aged 16 or more}}$
Households with 2 or more economically active persons	EM2ACT	$\frac{\text{Private households with 2 or more economically active persons}}{\text{Total private households with residents}}$
Unemployment	EMUNEM	$\frac{\text{All economically active persons aged 16 or more seeking work}}{\text{Total economically active persons aged 16 or more}}$
Students	EMSTU	$\frac{\text{Students}}{\text{Total persons aged 16 or more}}$
Married women at work	EMMWOM	$\frac{\text{Married women in employment}}{\text{Total married women in private households}}$
Unemployment among young people	EMUNYG	$\frac{\text{Residents aged 16-24 in private households economically active but out of employment}}{\text{Economically active residents aged 16-24 in private households}}$
Self-employed	EMSEMP	$\frac{\text{Self-employed persons}}{\text{Total economically active residents}}$

Table 5.2 (cont.)

Variable	Name	Calculation
6	ETHNICITY	
Born outside U.K.	ETNUKB	$\frac{\text{Persons not born in England, Scotland, Wales or rest of U.K.}}{\text{All residents}}$
New Commonwealth born persons	ETNOWP	$\frac{\text{Persons born in New Commonwealth or Pakistan}}{\text{All residents}}$
Households with New Commonwealth born head	ETNOWH	$\frac{\text{Total private households with New Commonwealth born head}}{\text{Total private households with residents}}$
7	MISCELLANEOUS VARIABLES	
Migrants	MMIG	$\frac{\text{Total residents aged over 1 year with a different address 1 year ago}}{\text{Total residents}}$
No car ownership	MNOCAR	$\frac{\text{Private households with no car}}{\text{Total private households with residents}}$

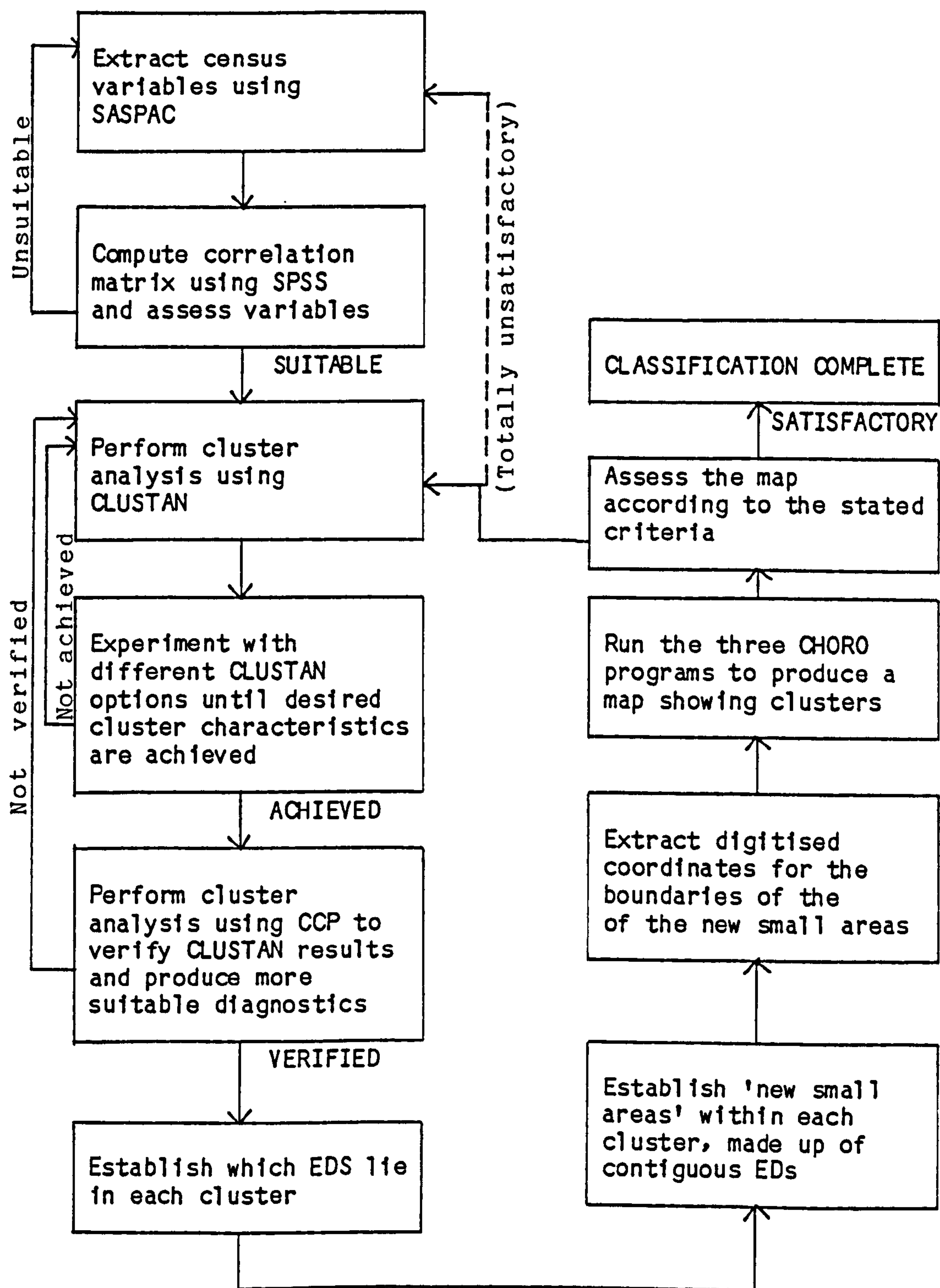
procedure would allow the greatest flexibility in testing alternative strategies. A method was devised, and relevant programs were written, which would enable most steps from the extraction of census data to the plotting of the final map to be done by computer. Figure 5.1 illustrates the processing sequence.

It should be noted at the outset that it was decided not to include a **contiguity constraint** within the clustering process. This was felt to be unnecessary for several reasons. First, contiguity *per se* can only be justified for cluster membership when spatial diffusion, and hence proximity, are key features of the process under study. In this project the main concern was not with the process of change within the urban area, nor was it felt that the short-term diffusion of socio-economic conditions should be the focus of interest. A second reason was the problem of 'chaining' which is exacerbated when a contiguity constraint is imposed. This results in a string of areas being allocated to a cluster, each related to the previous but successively having fewer characteristics in common with the first. Finally, it was felt that contiguity could more appropriately be introduced after the clustering process. In this way, the clusters were derived from the set of variables considered to be important, but the new small areas were delimited afterwards according to contiguity among cluster members.

Step 1. Extraction of variables using SASPAC

A major computer program, SASPAC, was designed specifically for handling the 1981 census small area statistics (LAMSAC,1980). Produced by the Universities of Durham and Edinburgh, and financed primarily by local authorities, SASPAC and the census data that can be processed by it have been made available to all universities through an agreement between ESRC and OPCS.

Figure 5.1 Classifying census enumeration districts - method



→ Arrows refer to the direction of processing.

The package provides various facilities within the overall objective of data extraction. These include an efficient system of data storage; the capacity to include or exclude census areas during processing through conditional selection; the ability to aggregate census areas into 'new zones' and produce variables for these; and a choice of output which includes pseudo-OPCS 'pages', lists of variables, or a 'matrix file' which may then be input to another package. SASPAC does not perform any statistical analyses, nor does it carry out mapping or other graphic displays of the results. (UMRCC,1982).

Step 2. Assessment of variables.

Table 5.2 listed the ratio variables suggested by the work of previous researchers to be appropriate for a general census classification. The first step of the analysis was to assess the usefulness of these variables collectively and to exclude those contributing little to the data set. A correlation matrix was drawn up and highly correlated pairs of variables were examined reducing the data set to a total of 34 variables (Table 5.3).

Step 3. Cluster analysis using CLUSTAN.

"Cluster analysis is an exploratory method for helping to solve classification problems... The object of a cluster analysis is to sort a sample of cases under consideration into groups such that the degree of association is high between members of the same group and low between members of different groups". (Wishart,1978, p.1).

Thus cluster analysis is the theoretician's answer to the relatively subjective process of classifying data. However, the results achieved are unique to the method and parameters used. For example, the same set of data may be classified using the procedures of hierarchic fusion or iterative relocation (see below for explanations of these) to give the same number of clusters, but the likelihood of these clusters containing identical cases is extremely low. Most users therefore consider cluster analysis to

Table 5.3 Variables used in the cluster analysis

DYNGCH	Young children
DOLDCH	Older children
DYNGWK	People of young working age
DMIDWK	People of middle working age
DOLDWK	People of older working age
DYNPNS	Young pensioners
DOLPNS	Older pensioners
HHSING	Single person households
HHCHIL	Households with many dependant children
HHLPAR	Lone parent households
HHXFAM	Households with children and 3+ adults
HSOWN	Owner-occupied households
HSCOUN	Council households
HSRENT	Privately rented households
HSAMEN	Households without exclusive use of amenities
HSCRWD	Overcrowded households
HSSHAR	Shared dwellings
HSFLAT	Flats
EMEACT	Economically active
EM2ACT	Households with 2 people economically active
EMUNEM	Unemployed
EMSTU	Students
EMMWOM	Married women at work
EMUNYG	Young unemployed
EMSEMP	Self-employed
ETNCWP	Persons born outside the UK
MMIG	Migrants
MNOCAR	Households with no car
SEPROF	Professional and managerial
SENMAN	Non-manual
SESMAN	Skilled manual
SESS	Semi-skilled manual
SEUS	Unskilled manual
SEEDUC	Educated

(See Table 5.2 for full specification of variables)

be a "descriptive and exploratory technique" (Openshaw, 1982d, p.5) and accept that the results obtained are only one of possibly several 'meaningful' sets, to be justified and explained in the context of the purpose of the classification.

Like SASPAC, the computer program CLUSTAN which performs cluster analysis was available at UMRCC and was run using data files held there. Various runs of CLUSTAN were performed on the census variables using different clustering algorithms and similarity coefficients. The number of component scores upon which the process was carried out was also varied. (see Appendix 4 for results of the PCA) The objective of these manipulations was to ensure that the results obtained from the classification process would be stable and robust under varying statistical conditions.

Two related algorithms were chosen. The first of these performed iterative relocation on the members of a user-defined number of clusters until some pre-defined function (in this case the error sum of squares-ESS) was optimised. This was repeated until a minimum number of clusters were achieved. The second algorithm suppressed the movement of cases from one cluster to another after the first classification had been produced, in effect simulating a hierarchical method. For the true iterative relocation a starting point of 20 clusters was requested, to be reduced eventually to two clusters. For the 'hierarchical' method 30 clusters were firstly identified and reduced again to two.

As shown in Table 5.4 there is a gradual increase in ESS as the number of clusters is reduced by **iterative relocation**. By examining the relative increase in ESS at each step, two possible 'break points' may be identified, perhaps indicating the 'natural' number of clusters for this data set. The first of these breaks occurs at the 16 cluster level, following which there is both a large increase in ESS (71.73) and a large decrease in the number

Table 5.4 Classification by iterative relocation: Run 1

Parameters:

- (i) 34 variables
- (ii) 7 component scores, each having eigenvalues greater than 1.0 following multivariate analysis without rotation.
- (iii) RELOCATE option 24, i.e. iterative relocation using an error sum of squares as a similarity coefficient. Objects are removed from the parent cluster for the relocation test.
- (iv) Starting number of clusters = 20
- (v) Finishing number of clusters = 2

Results:

No. of clusters	Error sum of squares (ESS)	Increase in ESS	No. of iterations	No. of relocations
20	1494.8611		12	1425
		27.13		
19	1521.9917		2	14
		36.09		
18	1558.0812		4	25
		45.92		
17	1604.0031		8	44
		37.60		
16	1641.6039		9	100
		71.73		
15	1713.3360		1	3
		72.84		
14	1786.1820		3	11
		68.99		
13	1855.1775		2	21
		56.88		
12	1912.0549		7	103
		74.82		
11	1986.8720		3	38
		85.23		
10	2072.1067		6	48
		103.50		
9	2175.6097		3	59
		144.94		
8	2320.5535		3	52
		195.56		
7	2516.1117		5	49
		294.79		
6	2810.9046		8	125
		307.41		
5	3118.3158		7	77
		378.86		
4	3497.1740		12	223
		516.15		
3	4013.3247		4	80
		751.56		
2	4764.8894		5	64

of relocations (from 100 to 3). This suggests that a temporarily stable solution has been found and that 16 clusters may well form a meaningful classification of these data.

A second break point occurs at 7 clusters which again is followed by a large increase in ESS (294.79 compared to 195.56 at the previous step). At this point however, the number of relocations also increases (from 49 to 125). This depends on both the nature of the cluster being split and on its similarity with other existing clusters. If 7 is a 'natural' number of clusters then the break-up of one of these will involve radical manipulation of the others in order to minimise ESS.

To test for the stability of this classification a separate procedure was set up. It was assumed that if the same final result (i.e. classification) could be achieved from different starting points (also classifications) then it would be likely that a global optimum had been achieved. The procedure is summarised in Table 5.5.

Two starting classifications were generated using iterative relocation and a third was assigned sequentially. Each set of 16 clusters was then reduced using the same algorithm and similarity coefficient to 2 final clusters. The results for the final 8 classifications were identical to those for Run 1, indicating that a stable classification was achieved. The convergence of the results prior to the 7 cluster grouping suggested this might be the most useful categorisation.

Given the interrelated nature of census data and the interdependence of many EDs, a hierarchical type of classification procedure may be considered intuitively to be more appropriate. Run 3 (Table 5.6) shows the results of such a procedure. The original 30 clusters were reduced to 2 on the basis

Table 5.5. Classification by iterative relocation: Run 2

Parameters:

- Step 1 RELOCATE with similarity coefficient 29 (16 clusters only).
 The result of this is stored for use as a starting classification for step 3.
- Step 2 RELOCATE with similarity coefficient 30 (16 clusters only).
 The result of this is stored for use as a starting classification for step 4.
- Step 3 RELOCATE with similarity coefficient 24. Reducing from the 16 clusters defined in step 1 to 2 final clusters.
- Step 4 RELOCATE with similarity coefficient 24. Reducing from the 16 clusters defined in step 2 to 2 final clusters.
- Step 5 RELOCATE with similarity coefficient 24. Reducing from the 16 sequentially allocated clusters normally used by RELOCATE as a starting point to 2 final clusters.

Results:

No. of clusters	Error sum of squares		
	Step 3	Step 4	Step 5

16	1649.2458	1663.6869	1720.2761
15	1696.3184	1706.5816	1749.6257
14	1768.1972	1774.5630	1792.4630
13	1840.1461	1831.5308	1853.8583
12	1912.2879	1912.0549	1930.6538
11	1996.5652	1986.8720	2005.9773
10	2072.2348	2072.1067	2082.0438
9	2175.6097	2175.6097	2175.6097
8	2320.5535	2320.5535	2320.5535
7	2516.1117	2516.1117	2516.1117
6	2810.9046	2810.9046	2810.9046
5	3118.3158	3118.3158	3118.3158
4	3497.1740	3497.1740	3497.1740
3	4013.3247	4013.3247	4013.3247
2	4764.8894	4764.8894	4764.8894

Table 5.6. Classification by hierarchic fusion: Run 3

No. of clusters	Clusters fused		Similarity Actual	coefficient Increase

30				
29	8	19	17.709	
28	1	20	19.319	1.61
27	24	27	21.680	2.361
26	8	23	28.286	6.606
25	2	12	28.480	0.194
24	25	28	28.771	0.291
23	3	16	30.006	1.235
22	13	30	37.613	7.607
21	1	26	38.499	0.886
20	9	21	43.085	4.586
19	6	22	52.948	9.863
18	7	8	54.842	1.894
17	3	5	61.451	6.609
16	2	17	64.141	2.69
15	11	18	67.958	3.817
14	14	15	70.226	2.268
13	4	29	77.134	6.908
12	6	24	78.659	1.525
11	1	9	85.363	6.704
10	10	13	97.692	12.329
9	4	7	122.826	25.134
8	6	25	163.753	40.927
7	11	14	174.474	10.721
6	3	6	270.319	96.345
5	2	10	280.558	9.739
4	2	11	373.169	92.611
3	2	4	566.496	193.327
2	1	2	747.793	181.297

of a minimal similarity coefficient.

By the very nature of the algorithm, the two most similar clusters will always be fused i.e. those with the lowest similarity coefficient. It does not, however, follow that the increase in the similarity coefficient always rises as it does in the relocation procedure. Between 30 and 12 clusters the table shows a gradual increase in the similarity coefficient, at each step rising by between 1 and 10; after this it fluctuates considerably, increasing by 96.345 after 7 clusters, before steadying and increasing again around 5 and 4 clusters. These results indicate that 7 clusters may again be the most appropriate number for this census data set.

Step 4. Cluster analysis using the Census Classification Program.

In this study the main reason for running Openshaw's (1982d) Census Classification Program (CCP) was to obtain a set of diagnostics more appropriate to census analysis. Although the output from procedure RESULT in CLUSTAN can provide a wide range of descriptive statistics, the statistical distributions of many census variables render these inappropriate; for example, cluster means and standard deviations are not particularly useful. Openshaw (1982d) lists other problems which his program overcomes, in particular the typically large volume of census data (CCP is designed to be computationally efficient) and the complexity of existing clustering packages. (CLUSTAN, for example, has no less than 40 types of similarity coefficient for different types of data, and numerous other options which may be selected for the different procedures).

The CCP comprises five separate programs which are run in a sequential fashion in batch mode. Normally the first program (CLUSTER1) is run once for any given data set and the resulting data file kept for later use. CLUSTER2, CLUSTER3 and CLUSTER4 may

be repeated until a meaningful classification is achieved; while CLUSTER5, which produces a large amount of printed output, is run only when the previous steps have been completed satisfactorily. The five programs perform the following computations:

CLUSTER1 : orthonormalises the data
CLUSTER2 : generates a starting classification (random or sequential)
CLUSTER3 : performs iterative relocation
CLUSTER4 : computes cluster diagnostics
CLUSTER5 : performs cluster evaluation
(Openshaw, 1982d p.9).

As the main reason for running the CCP was to confirm the CLUSTAN findings as described above, the objective was to replicate as far as possible, the steps taken previously in running CLUSTAN. Although the basic algorithm in the CCP is exactly the same as that used in the CLUSTAN option 'RELOCATE', the means of reducing the number of clusters is by hierarchic fusion only. It was possible therefore to replicate only the hierarchic reduction from 30 to 7 clusters using the CCP. To achieve the same final classification for iterative relocation from 20 to 7 clusters it was necessary to take the CLUSTAN output classification and input this directly to CCP, running only programs 1, 4 and 5.

The diagnostics produced fall into two broad groups. The first consists of cluster means and standard deviations which are presented along with relevant global values where appropriate. These figures may be interpreted either on their own, describing the cluster in terms of the absolute values of variable means or standard deviations, or in comparison with each other, by identifying variables in each cluster with either a mean value substantially different from the global mean value, or with a

standard deviation significantly smaller than the global equivalent. A sevenfold categorisation of 'z-scores' is also produced, which highlights variables according to how many standard deviations (SDs) the cluster mean is above or below the corresponding global mean. Table (a) of the 'key diagnostics' shows these results. It should be noted that 'z-scores' (or 'numbers of SDs') in this context have no statistical significance since the data are unlikely to be normally distributed. Table (b) of the 'key diagnostics' lists those variables in each cluster with standard deviations four or five times smaller than the global values, information which can be used to highlight potentially good diagnostic variables.

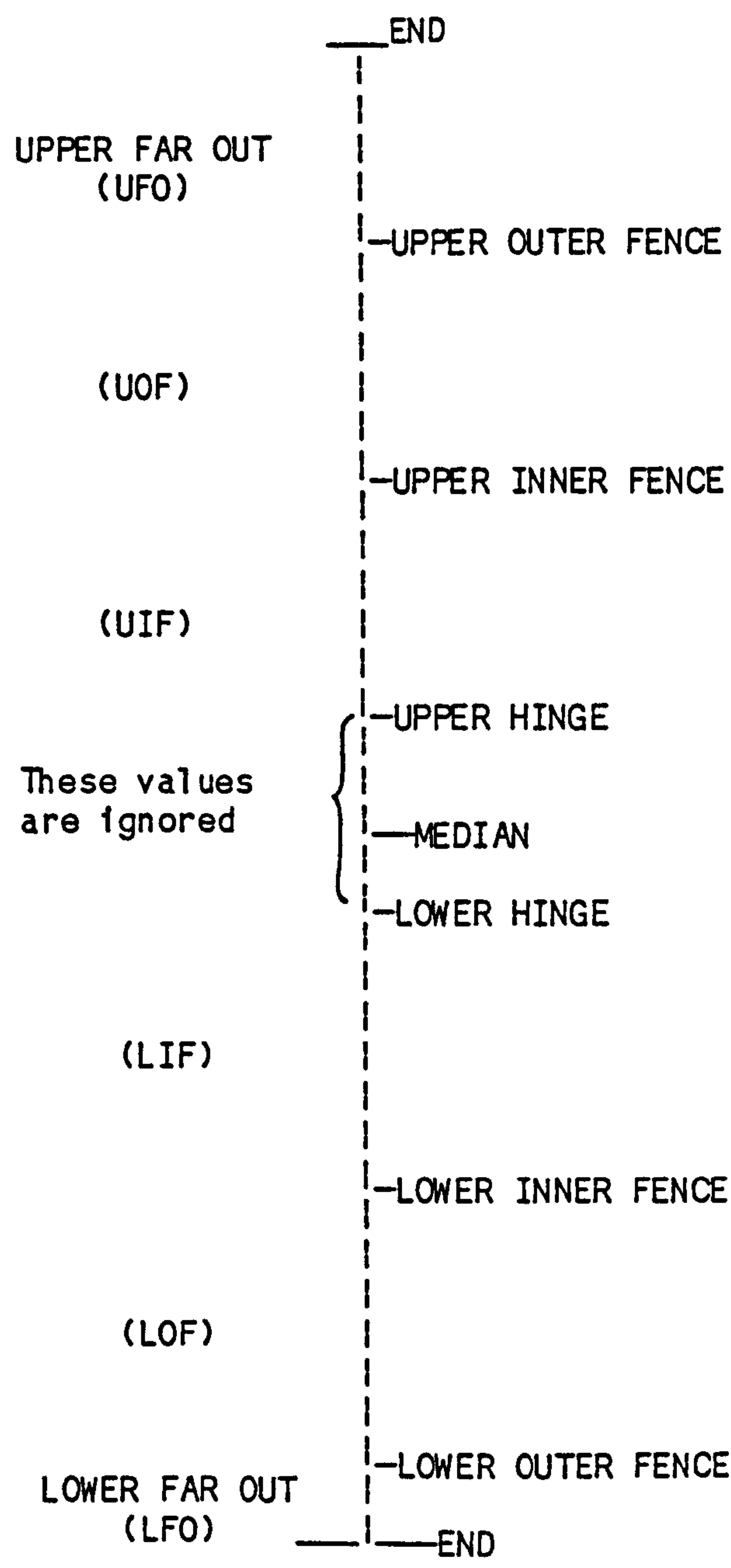
The second set of diagnostic statistics produced by the CCP are Tukey's (1977) 'robust distribution free measures' (Openshaw, 1982d, p.31) based on the median value. Openshaw describes these as follows:

"The median is simply the middle value of a ranked set of values. Ends are the smallest and largest values. Hinges are those values that lie half way between the ends and the median; they may also be called quartiles. Tukey (1977, p.44) now defines a rule of thumb to categorise values outside the two sets of hinges. A step is arbitrarily defined as 1.5 times the difference in values between the two hinges, an inner fence is set at one step outside the hinges, an outer fence is set at two steps outside the hinges, and values that lie beyond the outer fences are classed as being far out" (Openshaw, 1982, p.31).

Figure 5.2 is a diagrammatic representation of these summary statistics. It is of course possible that values for any one variable will not be found in all of these categories.

The main advantage of this type of statistic is that it does not make assumptions about the nature and frequency distribution of the data. As before, the objective is to identify for each cluster variables which are different from the global median. This is achieved by counting the number of cluster members in each of the categories between the hinges and the ends (i.e. 'inner

Figure 5.2 Tukey diagnostics



fence', 'outer fence' and 'far out' values, both above and below the global median). These figures are displayed as a percentage of all cases in the cluster and a 'cumulative' figure is also given. The cumulative count is a sum of cases outside the hinge area in either the lower or upper categories. For example, cluster 4 of the 7 cluster classification (see Table 5.7 (c)) has a cumulative percentage of 98.13 for variable ETNCWP (persons born in the New Commonwealth or Pakistan); of this total figure, 5.61% of cases lie in the 'upper inner fence' category, 27.1% in the 'upper outer fence', and 65.42% in the 'upper far out' category. These figures indicate that this variable has considerable diagnostic significance for this cluster. In order to extract only the most meaningful set of diagnostic variables table (c) of the 'key diagnostics' shows only those variables with a cumulative non-hinge count of 60% or more.

The aim of these diagnostics is to provide alternative descriptions of each cluster and thereby facilitate cluster labelling.

The key diagnostics for the original classification derived through iterative relocation are shown in Table 5.7. Based on an examination of the diagnostic statistics the following labels were chosen.

Cluster 1 is the largest of the seven clusters, it has no variables with cluster means significantly different from the global mean and there is no correspondence between the other diagnostic statistics. This indistinct cluster will be further investigated later.

Areas in Cluster 2 are clearly dominated by council accommodation and indicators of low social status. Owner-occupation (HSOWN) and privately rented property (HSRENT) are heavily represented in the 'lower inner fence' category and,

Table 5.7. Key diagnostics for 7 clusters derived through iterative relocation from 20 clusters

(a) Variables with cluster means significantly different from global mean.

Cluster	Size	Variables	No. of global SDs from global mean
1	220	None	
2	141	HSCOUN HSOWN	+ 1 - 1
3	35	HSAMEN, SENMAN, SEEDUC HHSING, HSRENT, MMIG, DYNGWK HSSHAR DOLDCH, EM2ACT	+ 1 - 2 + 3 - 1
4	107	HHCHIL, HHXFAM, HSAMEN, HSCRWD, EMUNEM ETNCWP	+ 1 - 2
5	167	EMEACT, EM2ACT, DMIDWK MNOCAR, DYNPNS	+ 1 - 1
6	45	HHSING, HHLPAR, HSCOUN, MNOCAR, DYNPNS HSFLAT, DOLPNS DOLDCH, HHXFAM, HSOWN, EMEACT, EM2ACT, DMIDWK	+ 1 + 2 - 1
7	132	EMSEMP, SEPROF, SEEDUC	+ 1

(b) Variables with cluster standard deviations significantly smaller than global standard deviations

Cluster	Size	Variables	Global SD = cluster SD X
1	220	HSSHAR	4
2	141	HSRENT, HSSHAR, SEEDUC, HSOWN SEPROF	5 4
3	35	HSCOUN	4
4	107	HSFLAT HSCOUN	5 4
5	167	HSRENT, HSAMEN, HSSHAR MNOCAR, HHSING, HHLPAR, HSOWN, HSCOUN, HSCRWD, HSFLAT, EMUNEM, SEUS, DOLPNS	5 4
6	45	HSAMEN HSOWN, HHCHIL, HSCRWD	5 4
7	132	HSCOUN, EMUNEM HHLPAR, HSOWN, HSAMEN, HSCRWD, ETNCWP	5 4

Table 5.7 (Cont.)

(c) Tukey style diagnostics: variables having a large percentage of cases in or beyond the 'inner fence'.

		Percentage in each category							
Cluster	Variable	LFO	LOF	LIF	CUM*	UFO	UOF	UIF	CUM*
1	HHCHIL			66.36	66.36				
	HSCRWD			64.09	64.09				
	HSFLAT			67.27	67.27				
	EMSTU			61.36	61.36				
2	HHLPAR							53.90	65.96
	HSCOUN							96.45	96.45
	HSOWN			97.16	97.16				
	HSRENT			86.52	86.52				
	EMSTU			73.05	73.05				
	EMSEMP			78.01	78.01				
	SEPROF			68.79	68.79				
	SEEDUC			68.79	68.79				
	DMIDWK			66.67	66.67				
3	HHSING					48.57	34.29	94.29	
	HSRENT					51.43	8.57	91.43	
	MMIG					34.29	31.43	97.14	
	DYNGWK					51.43	17.14	77.14	
	HSAMEN						57.14	80.00	
	SENMAN						57.14	65.71	
	SEEDUC						48.57	65.71	
	DYNGCH			60.00	60.00				
	DOLDCH			68.57	88.57				
	HHXFAM			71.43	71.43				
	EM2ACT			74.29	80.00				
4	ETNCWP					65.42	27.10	5.61	98.13
	HSCRWD					39.25	20.56	95.33	
	HHCHIL						38.32	66.36	
	HHXFAM						48.60	84.11	
	HSRENT						63.55	71.03	
	HSAMEN						70.09	71.96	
	EMUNEM						65.42	78.50	
	HSFLAT			85.98	85.98				
	DOLDWK			63.55	63.55				
5	EMEACT							71.26	71.86
	EM2ACT							72.46	73.05
	EMMWOM							64.07	64.07
	DMIDWK							41.32	65.27
	HHSING			73.05	73.05				
	HHCHIL			70.06	70.06				
	HSCRWD			77.84	77.84				
	HSFLAT			79.64	79.64				
	EMUNEM			74.25	74.25				
	MNOCAR			78.44	78.44				
	SEUS			76.65	76.65				
	DYNPNS			70.66	70.66				
	DOLPNS			80.84	80.84				

CUM = cumulative percentage in upper or lower category.

/Continued...

Table 5.7 (Cont.)

		Percentage in each category							
Cluster	Variable	LFO	LOF	LIF	CUM*	UFO	UOF	UIF	CUM*

6	HSFLAT					73.33	2.22	0.0	75.56
	HHSING							48.89	95.56
	HHLPAR							31.11	62.22
	HSCOUN							91.11	91.11
	MNOCAR							84.44	84.44
	DYNPNS							75.56	86.67
	DOLPNS							33.33	82.22
	DYNGCH			62.22	62.22				
	DOLDCH			62.22	80.00				
	HHCHIL			91.11	91.11				
	HHXFAM			88.89	88.89				
	HSOWN			88.89	88.89				
	HSRENT			68.89	68.89				
	HSCRWD			77.78	77.78				
	EMEACT			57.78	86.67				
	EM2ACT			86.67	97.78				
	EMSTU			91.11	91.11				
	DYNGWK			60.00	68.89				
	DMIDWK			57.78	86.67				
7	EMSEMP							53.30	72.73
	SEPROF							55.30	75.76
	SEEDUC							54.55	68.18
	HHCHIL			82.58	82.58				
	HSCOUN			62.88	62.88				
	HSCRWD			80.30	80.30				
	HSFLAT			75.76	75.76				
	EMUNEM			77.27	77.27				
	SESMAN			63.64	63.64				
	SEJS			84.85	84.85				

*Cumulative percentage in upper or lower category.

moreover, have cluster standard deviations five times smaller than the equivalent global values. Council housing (HSCOUN) in contrast, has 96.45% of cases in the 'upper inner fence' category and a cluster mean significantly higher than the global mean. Indicators of high socio-economic status such as professional occupations (SEPROF), post-school education (SEEDUC) and self-employment (EMSEMP) are under-represented while lone parent families (HHLPAR) are prominent. **"Low status, council"** was selected as the cluster name.

Members of Cluster 3 have particularly high proportions of single person households (HHSING), privately rented property (HSRENT), migrancy (MMIG) and young adults (DYNGWK). The cluster mean value for the variable 'shared dwellings' is over three times the global value. Indicators of potential residential stability such as children (DYNGCH and DOLDCH), mixed households (HHXFAM) and households with two or more people economically active (EMEACT) are under-represented. This cluster therefore includes areas with a predominantly young, probably upwardly mobile population living in rented and shared accommodation. It is labelled **"upwardly transient"**.

"Low status, ethnic quarter" best describes the areas in Cluster 4 since almost all EDs have values for ethnicity (ETNCWP) in the upper inner fence and above. Overcrowded households (HSCRWD), mixed households (HHXFAM), and those with many dependant children describe areas of high density living, other characteristics of these areas include households lacking amenities (HSAMEN) and high unemployment (EMUNEM).

Variables describing economic activity and the working population are important in describing Cluster 5 since it has high upper inner fence percentages for economically active persons

(EMEACT), households with two or more economically active (EM2ACT), married women at work (EMMWOM) and people in the 25 to 44 years age group (DMIDWK). The cluster has low deprivation indices - unemployment (EMUNEM), households with no car (MNOCAR), overcrowding (HSCRWD), large families (HHCHIL) and the unskilled occupational group (SEUS) - and people over pensionable age are under-represented (DYNPNS and DOLDPNS). A suitable label for these areas is therefore **"high social status, low density"**.

Cluster 6 is a small cluster and therefore has cluster mean values significantly different from the global mean for a large number of variables. Purpose-built flats is an important variable in forming the cluster since over 70% of cases lie in the 'upper far out' category. Above average mean values for the pensionable age groups, single person households (HHSING), households without a car (MNOCAR), council tenancy (HSCOUN) and lone parent families (HHLPAR) indicate that this could be a cluster of areas dominated by the **socially dependant** population living in council accommodation. Relatively low values for variables representing familism - children, younger adults and economic activity - and also for non-council tenancy, support this designation.

The final cluster is the only one to be mainly described by indicators of socio-economic status. With above average areal proportions for professional, self-employed and educated people (SEPROF, EMSEMP and SEEDUC) and correspondingly low values for skilled manual and unskilled occupational groups (SESMAN and SEUS) the cluster can be appropriately labelled **"high socio-economic status"**. The cluster has relatively few households with large numbers of children (HHCHIL) and has little overcrowding (HSCRWD). Council tenancy (HSCOUN) and purpose-built flats (HSFLAT) are under-represented.

Table 5.8 contains the key diagnostics for the second

Table 5.8. Key diagnostics for 7 clusters derived through hierarchic fusion from 30 clusters.

(a) Variables with cluster means significantly different from the global mean.

Cluster	Size	Variables	No. of global SDs from global mean
1	110	HHCHIL, HHXFAM, HSCRWD, EMUNEM ETNCWP	+ 1 + 2
2	129	None	
3	90	EMSEMP, SEPROF, SEEDUC	+ 1
4	69	HHSING, HSAMEN, MMIG, SEEDUC, DYNGWK DOLDCH HSRENT, HSSHAR	+ 1 - 1 - 2
5	256	MNOCAR	- 1
6	51	HHSING, HHLPAR, HSCOUN, MNOCAR, DYNPNS, -DOLPNS DOLDCH, HSOWN, EMEACT, EM2ACT, DMIDWK HSFLAT	+ 1 - 1 + 2
7	142	HSCOUN HSOWN	+ 1 - 1

(b) Variables with cluster standard deviations significantly smaller than global standard deviations.

Cluster	Size	Variables	Global SD = cluster SD X
1	110	HSCOUN, HSFLAT	5
2	129	HSSHAR HSCRWD, HSFLAT	5 4
3	90	HSOWN, HSCOUN HHLPAR, HSRENT, HSAMEN, HSSHAR	5 4
4	69	HSCOUN	4
5	256	HSCRWD, HHSAR, ETNCWP HHCHIL, HHLPAR, HSOWN, HSCOUN, HSRENT, HSAMEN, HSFLAT, EMUNEM	5 4
6	51	HSAMEN, HSSHAR HSOWN, HHCHIL, HSCRWD	5 4
7	142	HSOWN, HSRENT, HSSHAR, SEEDUC SEPROF	5 4

Table 5.8 (Cont.)

(c) Tukey style diagnostics: variables having a large percentage of cases in or beyond the 'inner fence'.

		Percentage in each category							
Cluster	Variable	LFO	LOF	LIF	CUM	UFO	UOF	UIF	CUM

1	ETNCWP					67.27	27.27	4.54	99.09
	HSCRWD						40.00	20.91	93.64
	HHCHIL							38.18	64.55
	HSRENT							60.91	67.27
	HSAMEN							67.27	69.99
	EMUNEM							61.82	73.64
	HSFLAT			86.36	86.36				
	DOLDWK			65.45	65.45				
2	HHCHIL			67.44	67.44				
	HSCRWD			65.89	65.89				
	HSFLAT			65.89	65.89				
	EMSTU			72.09	72.09				
3	EMSEMP							56.67	78.89
	SEPROF							50.00	74.44
	SEEDUC							56.67	67.78
	DYNGCH			62.22	62.22				
	HHCHIL			74.44	74.44				
	HSCOUN			66.67	66.67				
	HSCRWD			75.56	75.56				
	HSFLAT			74.44	74.44				
	EMEACT			66.67	66.67				
	EMUNEM			67.78	67.78				
	SESMAN			67.78	67.78				
	SEUS			82.22	82.22				
4	HSRENT					36.23	37.68	91.30	
	HHSING							50.72	88.41
	HSAMEN							49.28	60.87
	MMIG							47.83	89.86
	DYNGWK							34.78	66.67
	DOLDCH			69.57	82.61				
	HHCHIL			63.77	63.77				
	HHXFAM			71.01	71.01				
	EM2ACT			63.77	66.67				
5	HHCHIL			74.22	74.22				
	HSCRWD			80.86	80.86				
	HSFLAT			78.52	78.52				
	EMUNEM			71.48	71.48				
	MNOCAR			64.45	64.45				
	SEUS			75.78	75.78				

/Continued...

Table 5.8 (Cont.)

		Percentage in each category							
Cluster	Variable	LFO	LOF	LIF	CUM	UFO	UOF	UIF	CUM

6	HSFLAT					84.31	1.96	0.0	86.27
	HHSING							52.94	90.19
	HHLPAR							31.37	62.75
	HSCOUN							92.16	92.16
	MNOCAR							84.31	84.31
	DYNPNS							66.67	78.43
	DOLPNS							23.53	62.75
	DOLDCH			60.78	74.51				
	HHCHIL			94.12	94.12				
	HHXFAM			86.27	86.27				
	HSOWN			90.20	90.20				
	HSRENT			72.55	72.55				
	HSCRWD			70.59	70.59				
	EMEACT			45.10	70.59				
	EM2ACT			78.43	90.20				
	EMSTU			88.24	88.24				
	EMSEMP			60.78	60.78				
	DYNGWK			54.90	62.75				
	DMIDWK			52.94	78.43				
7	HHLPAR							54.93	64.08
	HSCOUN							97.18	97.18
	HSOWN			95.07	95.07				
	HSRENT			84.51	84.51				
	EMSTU			73.24	73.24				
	EMSEMP			76.76	76.76				
	SEPROF			66.2	66.20				
	SEEDUC			67.61	67.61				
	DMIDWK			68.31	68.31				

classification, that of 7 clusters derived by hierarchic fusion. It should be noted that the hierarchic process is less likely to give clear characteristics since it precludes the movement of cases from one cluster to another after the original fusion. As a result, once a case has been included in a cluster it must remain there, even if its characteristics might subsequently be more appropriate to another.

Cluster 1 is of a similar size and character to that labelled "low status ethnic quarter", in the previous classification. Ethnicity, overcrowding, many dependant children and unemployment all have above average cluster means.

The areas comprising Cluster 2 on the other hand, have no variables with exceptional cluster means, and only below normal percentages of students, overcrowded households, flats and large families. This could be tentatively described as a cluster representing 'low density living' but is perhaps more realistically labelled "indistinct".

Variables describing socio-economic status (EMSEMP, SEPROF, SEEDUC, SEUS and SESMAN) are important in Cluster 3. Like its counterpart in the previous classification, this "high socio-economic status" cluster is dominated by areas showing the characteristics of achievement and under-represented by those associated with deprivation.

Cluster 4 is the most easily labelled cluster, describing as it does areas frequented by the transient population. Rented and shared accommodation, single status and high migrancy rates are all clearly identified.

Cluster 5 is difficult to assess. In a less than satisfactory negative way it identifies "middle class" areas, since it is characterised by below average proportions of

unskilled headed households, those without a car, overcrowding, large families, unemployment and purpose-built flats. Its large size (256 areas) precludes any significant variation from global values.

The smallest cluster (Cluster 6 with 51 members) has many variables for each of the diagnostic statistics. Overall they describe areas of disadvantage, being dominated by flats, council housing, older people and lone parent families. This cluster is closest in character to that labelled "socially dependant, council" in the previous classification.

The final cluster is also similar in both size and character to one previously identified as "low status, council". Tenancy is the most significant dimension of the cluster, along with variables closely associated with council households.

Despite their different starting points and clustering algorithms, there are strong similarities between the two classifications. This points to stability within the data set and suggests that it might be possible to experiment with different sets of seven clusters in order to achieve an optimum set for the data. This should be measured in terms of average cluster size, ease of labelling and intuitive reasoning. Table 5.9 summarises the results of seven such attempts at classifying the data set. The first of these used a starting classification derived through sequential assignment; the next five had random starting classifications, each beginning with a different random number; the final column shows equivalent figures for the classification described above in which iterative relocation was used successively on a series of classifications to reduce from 20 to 7 clusters. Table 5.10 shows cluster sizes and labels for each of the seven classifications. Cluster numbers in this table have no significance.

Table 5.9 Identifying a 'better' set of clusters - diagnostic statistics

	Sequential Assignment	Random Starting Classifications ←——beginning with sequence number→					Relocation from 20 to 7 clusters
		100	200	300	400	500	
No. of iterations	14	12	7	13	13	10	-
No. of moves	1288	1139	970	1186	1277	1140	-
Final within cluster sum of squares	2519.68	2613.18	2612.46	2521.76	2522.72	2519.71	2516.11
Mean within cluster distance to centroid	3.6886	3.0077	3.0048	2.9461	2.9486	2.9475	3.6942
Mean between cluster distance to centroid	7.3207	7.1302	7.011	6.9119	7.0252	7.0215	7.4206
Ratio of within to between centroid distances	0.5038	0.4218	0.4286	0.4262	0.4197	0.4198	0.4978
Mean distance between cluster centroids	6.7188	6.8903	6.6849	6.7076	6.8989	6.8716	6.8946

Table 5.10 Identifying a 'better' set of clusters - cluster labelling

(i) Original classification assigned sequentially.

		SIZE
1	Indistinct	210
2	Low status, council	142
3	Transient	51
4	Low status, ethnic quarter	107
5	High status, low density	165
6	Socially dependant, council	41
7	High income areas	131

(ii) Original classification assigned randomly, starting at the 100th random number.

1	Indistinct	245
2	Elderly population	47
3	Established council	106
4	Newer council	66
5	Middle income areas	242
6	Transient	33
7	Low status, ethnic quarter	108

(iii) Original classification assigned randomly, starting at the 200th random number.

1	Middle income areas	235
2	Low status, council	106
3	Socially dependant, council	65
4	Low status, ethnic quarter	107
5	Transient	53
6	Indistinct	242
7	Elderly population	39

(iv) Original classification assigned randomly, starting at the 300th random number.

1	Low status, council	142
2	Elderly population	41
3	Transient	51
4	High status, low density	128
5	Low status, ethnic quarter	108
6	Indistinct	216
7	High income areas	161

/Continued...

Table 5.10 (Cont.)

(v) Original classification assigned randomly, starting at the 400th random number.

		SIZE
1	Low status, ethnic quarter	108
2	Indistinct	223
3	Low status, council	142
4	Socially dependant, council	49
5	Transient	33
6	Middle class, low density	159
7	High status, low density	133

(vi) Original classification assigned randomly, starting at the 500th random number.

1	Transient	33
2	Low status, ethnic quarter	107
3	High status, low density	167
4	Elderly population	49
5	High socio-economic status	132
6	Indistinct	217
7	Low status, council	142

(vii) Original classification of 20 clusters assigned sequentially, reduced to 7 clusters through repeated iterative relocation.

1	Indistinct	220
2	Low status, council	141
3	Upwardly transient	35
4	Low status, ethnic quarter	107
5	High status, low density	167
6	Socially dependant, council	45
7	High socio-economic status	132

Several comments arise from the cluster labelling process. First, and most important, the labels give a general description to each cluster, they do not accurately represent every ED in the cluster. Secondly, larger clusters are on the whole less easy to label, possibly because the larger number of cases produce mean values closer to the global mean. Thirdly, some clusters, such as "low status ethnic quarter" were much easier to label than others, such as "middle income". Fourthly, in comparing the classifications it should not be assumed that those given the same name are identical within since they vary considerably in size and composition. Finally, and also very important, clusters of equal size and with similar characteristics do not necessarily contain the same assemblage of EDs.

The diagnostic statistics were next used to compare the seven classifications. As a general rule, a preferred classification is one that both minimises the final within cluster sum of squares, the within cluster distance to centroid (these are roughly equivalent measures), the ratio of within to between centroid distances and the mean distance between cluster centroids, and maximises the mean between cluster distance to centroid. Based on these premises, the fourth and final classifications appeared to be the 'best'. The random starting classification beginning at the 300th sequence number resulted in the smallest within cluster distance to centroid (2.9461) and the smallest mean distance between cluster centroids (6.7076). That derived by iterative relocation from 20 to 7 clusters minimises the within cluster sum of squares (2516.11) and maximises the mean between cluster distance (7.4206). The random starting classification from the 400th sequence number produces the smallest ratio of within to between cluster distances.

The cluster labels for the seven classifications suggest the following common characteristics:

- (i) **Low status ethnic quarter** - a consistent cluster throughout.
- (ii) **Low status, council** - with the exception of classifications 2 and 3 this is also a stable cluster.
- (iii) **Transient** - this is present in all classifications but varies in size, perhaps losing or gaining areas to the 'socially dependant, council' category.
- (iv) **Socially dependant, council** - this cluster can be identified in each classification but having several important dimensions it is difficult to give it a simple label.
- (v) **High socio-economic status** - this cluster appears in different guises across all seven classifications. In all it is associated with 'middle class suburbia' but it is rather nebulous because there is no direct measurement of 'middle class' status in the data set.
- (vi) **Indistinct** - a miscellaneous collection of EDs.

Because they do not perform well on the diagnostic statistics and are not representative of the general set of clusters described above, classifications 2 and 3 (random starts from the 100th and 200th random number respectively) are rejected. Classifications 1, 4 and 5 (sequential and random starts from the 300th and 400th sequence number) must also be viewed with some caution since they each include a 'high income' or 'middle class' cluster of significant size. Although acceptable in concept, such groups are difficult to define quantitatively and perhaps are not as satisfactory for future analysis as the more directly labelled "high socio-economic status" cluster. Of the two classifications remaining, number 6 is satisfactory, but number 7 (derived by successive iterative relocations) performs the better on the

diagnostic statistics. The within cluster sum of squares, which it minimises, is arguably the most important statistics of all. Since it is also representative of the general pattern of clusters it is therefore chosen for further analysis.

The first important manipulation of the selected classification was to reduce the size of the 'indistinct' cluster. This was done by re-classifying its constituent EDs. Table 5.11 shows the results of a further CLUSTAN run, this time using only the 220 EDs from the 'indistinct' cluster as the data set. As before, a 'break point' was identified, between 3 and 2 clusters where there is a sudden increase in ESS. It is suggested therefore that 3 clusters forms a useful division of this data set. Table 5.12 shows the key diagnostics for the new clusters (Clusters 1, 8 and 9).

Once again there is a 'residual' cluster which is difficult to label. Although the 'key diagnostics' suggest very little, the raw data (see Table 5.13) indicate that Cluster 1 comprises areas with above average owner-occupation, skilled manual workers and younger pensioners. In conjunction with the below normal levels of overcrowding, large families and flats, a name such as **"established private residential"** might be appropriate.

Nearly all the variables describing areas in Cluster 8 are direct or surrogate indicators of life cycle stage, ie. households with several children, extended families, students, young people, and more generally, those that are economically active. Younger pensioners are the only group to have above average representation in this cluster, hence the label **'elderly population'**.

Accommodation appears to be the leading dimension of Cluster 9, with 81.43% of EDs in the upper, non-fence categories for privately rented accommodation. Over 60% of EDs live in the same

Table 5.11. Reclassifying the enumeration districts in Cluster 1 (the 'indistinct' cluster).

Parameters:

- (i) 34 variables
- (ii) 220 enumeration districts
- (iii) 7 component scores
- (iv) RELOCATE option 24
- (v) Starting number of clusters = 15
- (vi) Finishing number of clusters = 2

Results:

No. of clusters	Error sum of squares (ESS)	Increase in ESS	No. of iterations	No. of relocations
15	497.5959		5	283
		8.46		
14	506.0561		4	8
		10.99		
13	517.0472		3	19
		16.04		
12	533.0893		3	15
		12.39		
11	545.4796		7	32
		21.73		
10	567.2083		3	10
		22.56		
9	589.7727		5	39
		25.44		
8	615.2157		3	20
		50.5		
7	665.7164		4	11
		45.63		
6	711.3464		3	25
		59.55		
5	770.9003		3	15
		64.46		
4	835.3624		3	11
		70.95		
3	906.3166		5	28
		120.19		
2	1026.5097		5	29

Table 5.12 Key diagnostics for three new clusters.

(a) Variables with cluster means significantly different from global mean.

Cluster	Size	Variables	No. of global SDs from global mean
1	97	None	
8	53	DYNPNS	+ 1
9	70	HSRENT	+ 1

(b) Variables with cluster standard deviations significantly smaller than global standard deviations.

Cluster	Size	Variables	Global SD = cluster SD X
1	97	HSSHAR, EMUNEM	5
		HSCRWD, HSFLAT, EM2ACT, ETNCWP	4
8	53	HSSHAR	5
		HHCHIL, HSCRWD, EMUNEM, EMSTU, ETNCWP	4
9	70	HSCOUN	5
		HSFLAT, MNOCAR, DOLPNS	4

(c) Tukey-style diagnostics: variables having a large percentage of cases in or beyond the 'inner fence'.

		Percentage in each category						
Cluster	Variable	LFO	LOF	LIF	CUM	UFO	UOF	UIF CUM
1	HHCHIL			65.98	65.98			
	HSCRWD			74.23	74.23			
	HSFLAT			69.07	69.07			
8	DYNPNS						66.04	66.81
	HHCHIL			81.13	81.13			
	HHXFAM			73.58	73.58			
	HSCRWD			79.25	79.25			
	EMEACT			67.92	67.92			
	EM2ACT			71.7	71.70			
	EMSTU			96.23	96.23			
	DYNGWK			67.92	67.92			
9	HSRENT						68.57	81.43
	HSAMEN						60.00	61.43
	HSFLAT			81.43	81.43			

**Table 5.13 Cluster 1. Variables with a cluster mean greater than
the global mean**

		Cluster mean	Global mean
~~~~~			
Owner-occupied households	(HSOWN)	75.1649	60.49
Skilled manual s.e.g.	(SESMAN)	36.433	29.4711
'Young' pensioners (to age 74)	(DYNPNS)	14.0515	11.3436
(Married women at work	(EMMWOM)	52.0722	50.2184)
(self-employed people	(EMSEMP)	7.7526	7.5608)
(People aged 45 to retirement)	(DOLDWK)	20.5464	19.2562)
(People aged 75 and over	(DOLPNS)	6.7423	5.8796)

categories for lacking exclusive use of all amenities. Below the lower inner fence 81.43% of EDs have relatively few purpose-built flats. Several other variables have cluster means larger than the global means. The most significant of these include single person households, owner-occupiers, New Commonwealth and Pakistan-born people, households without a car and young adults. A suitable label for this cluster is areas of **private renting**.

It has been found therefore, that a 'derived' set of 9 clusters best describes this census data set. Although originally a set of 7 clusters was identified, the division of the largest and least useful of these into three smaller clusters improved the performance of the classification both in terms of its diagnostic statistics and ease of cluster labelling.

#### **Step 5. Processing the output from the classification.**

CLUSTAN produces as output a list of case numbers contained in each of the clusters identified, the first stage of further analysis was therefore to give these their respective ED names and to examine their relative location.

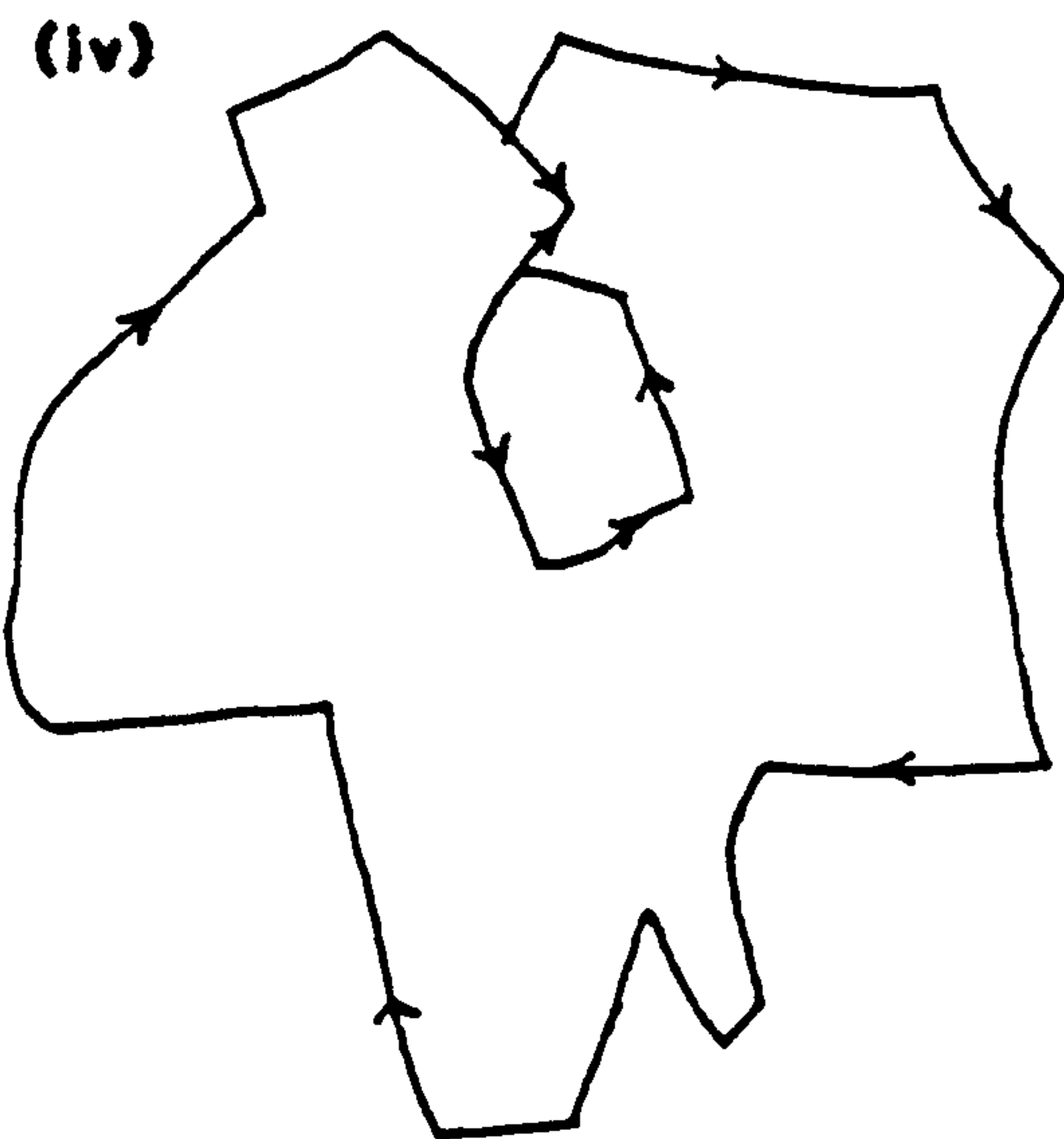
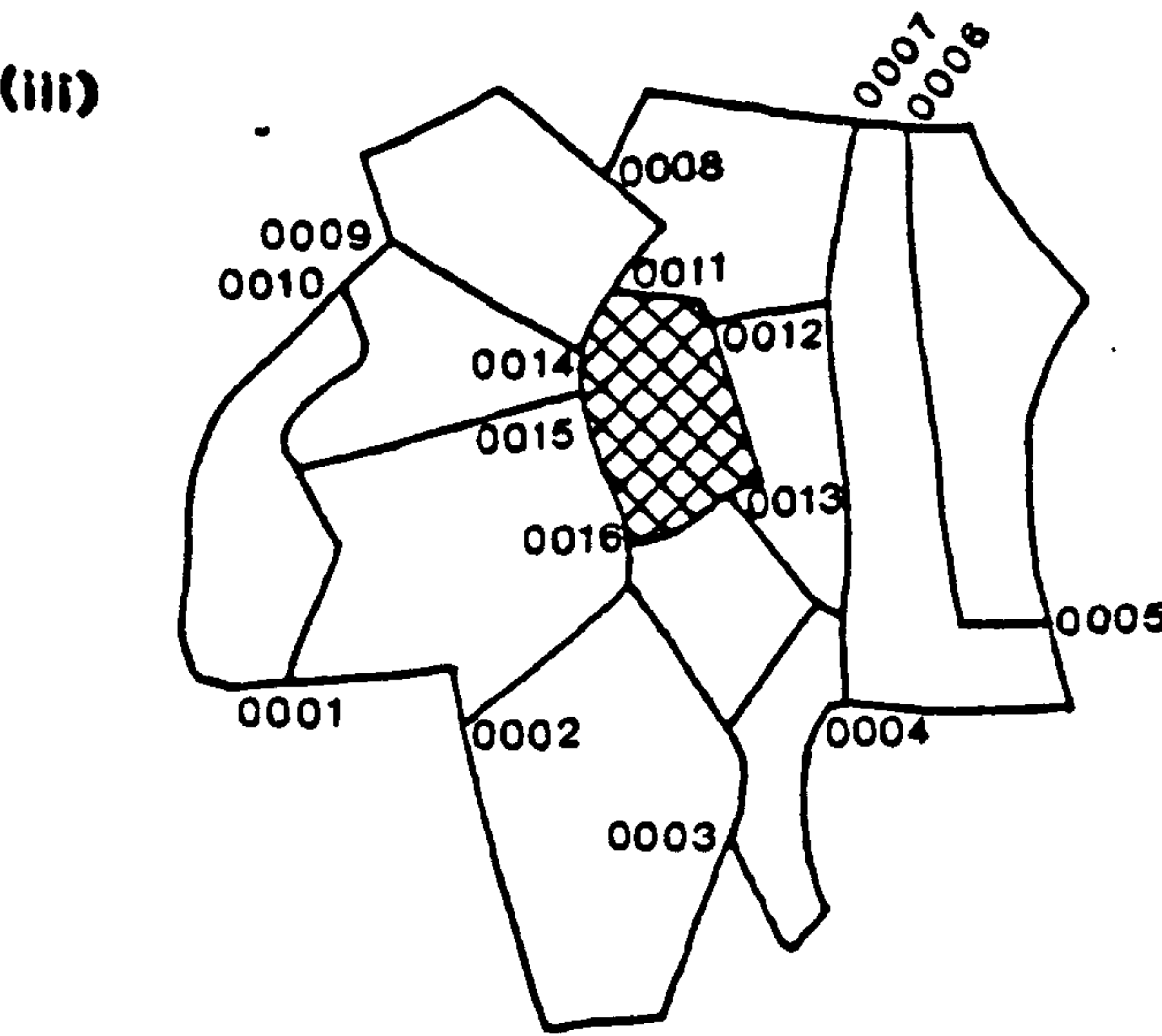
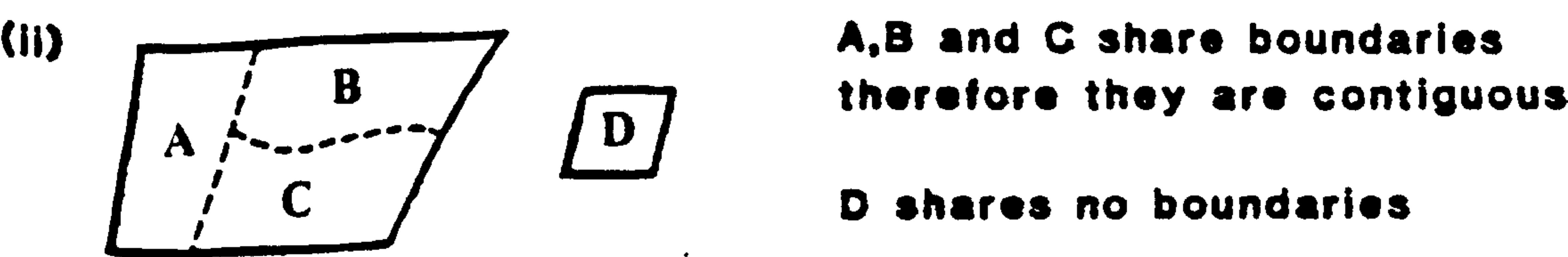
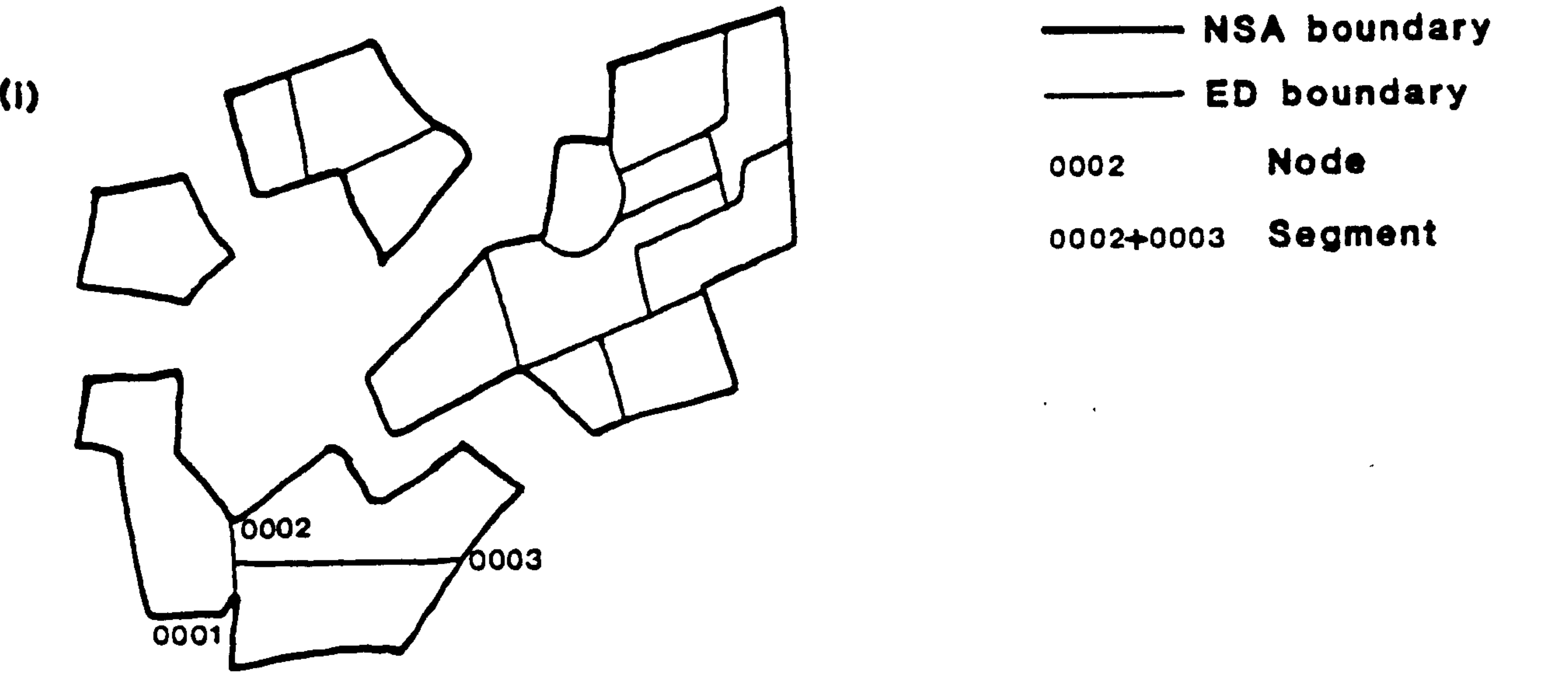
For practical and methodological reasons it had already been decided that the cluster analysis should not include as a variable a contiguity constraint, so EDs in each cluster are scattered across the study area, sometimes contiguous, sometimes not. A computer program, MPAGGR (see Appendix 2.2) was written to determine which EDs were connected and to redefine the boundaries of these into larger 'new small areas' (NSAs). The original definition of the ED boundaries was important to this program and is therefore described below.

The boundary of each ED was described by a sequence of 'segments', in turn defined as a link between two four digit 'nodes' (see Figure 5.3(i)). The algorithm used by MPAGGR to determine which EDs were contiguous simply examined a list of



**Figure 5.3**

**Creating new small areas from enumeration district boundaries**



segments and where common segments were found between EDs they were assumed to be contiguous (see Figure 5.3(ii)). Special processing was introduced to cater for the case where several EDs from one cluster completely surrounded an ED or EDs from another cluster (see Figure 5.3(iii)). Program MPHOLE (Appendix 2.2) used a 'point-in-polygon' algorithm to test for this condition. If found, a message was printed to state which area lay within which other area, the file containing NSA boundaries was then adjusted manually. In the above example, the redefined NSA would be described as follows:

```
(START)
0001+0010
0010+0009
0009+0008
0008+0011
0011+0014
0014+0015
0015+0016
0016+0013
0013+0012
0012+0011
0008+0011
0008+0007
0007+0006
0006+0005
0005+0004
0004+0003
0003+0002
0002+0001
(END)
```

(see Figure 5.3(iv)). As shown, the boundary of the smaller area is incorporated into the larger by linking the areas with a 'bridge'. This was necessary for ease of future programming since the mapping package used did not recognise 'islands' unless they were identified in this manner.

The output from this procedure was a list of NSAs defined by their boundary segments. Coordinates for these were then extracted from a base file containing digitised coordinates for all segments in Greater Leicester, and this formed the new base file for input to the next stage of processing.

CHORO, written by Blakemore and Reeve (1977) is a package for drawing choropleth maps on a graph plotter which has been extensively modified by the University of Leicester Computer Centre. The package is run in three stages, although after the first and second have been completed successfully it is possible to produce a series of maps by just altering variables in the control file of the third.

The first program is the **segment definition** program which reads a file of digitised segments (such as that defining NSA boundaries) and creates two random access files. The second, the **area definition** program, reads the segment definition file just produced, and a second file containing a list of area names and the segments which make up those areas. For the original Leicester data, the area definition file input to this program was produced by another program, MPPOLX (Appendix 2.2). This read the digitiser file containing coordinates and extracted the smallest polygons created by the segments ie. EDs. It ensured that each ED was defined by the segments surrounding it specified in a clockwise direction. The major advantages of using a second program to produce an input file of this nature were that it assumed 'clean' source data and it eliminated another potential for human error.

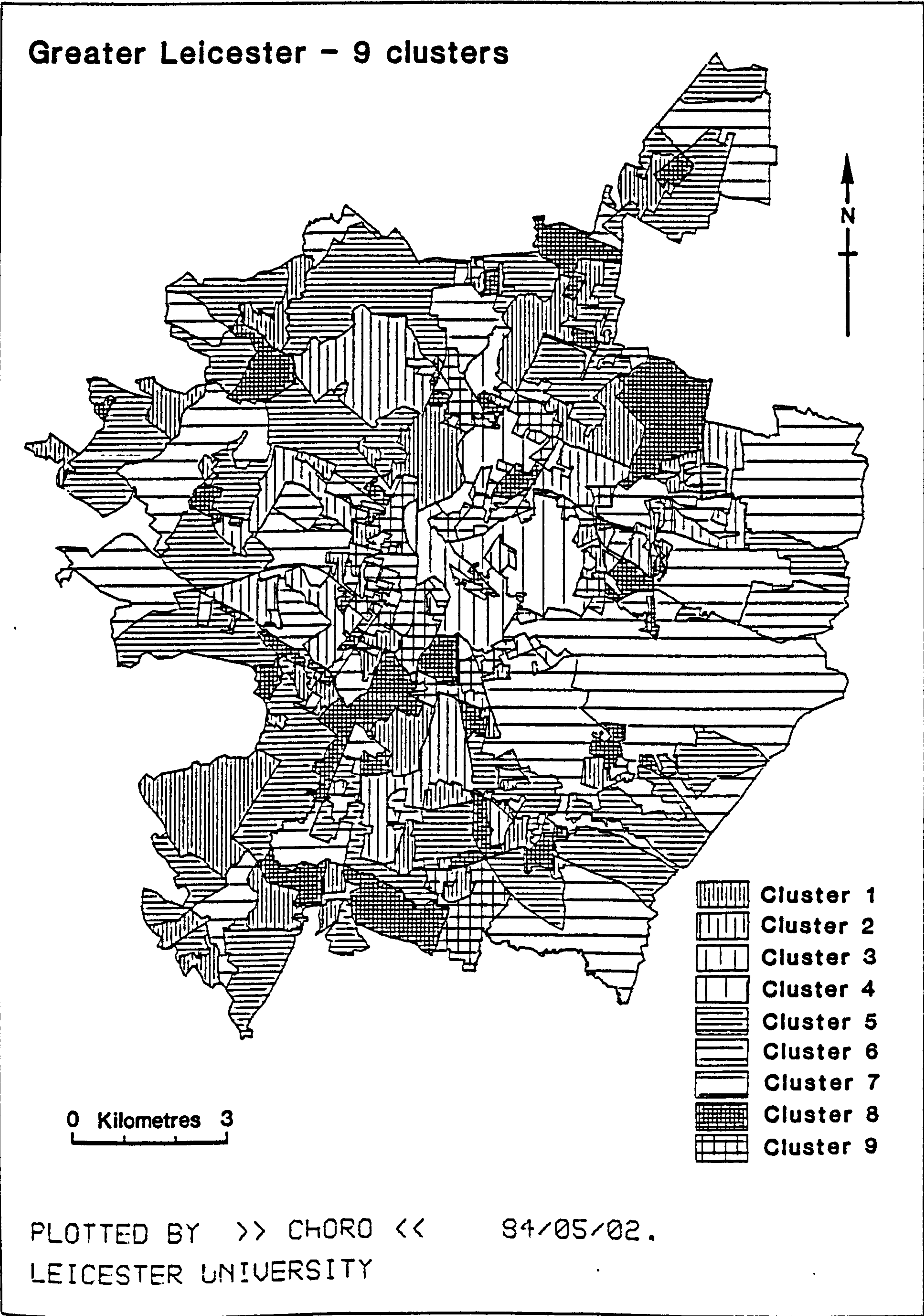
The last stage of CHORO, the **drawing program**, uses the random access files created in area and segment definition and a control file which determines the form in which the maps are to appear. The file 'MPFDW' (Appendix 2.3) contains an example of the control commands used to draw a single map, and Figure 5.4, showing all 9 clusters in Greater Leicester, is typical of the output.

#### **Step 6. Assessment of the classification.**

The number of NSAs is important for ease of subsequent



Figure 5.4



processing; 200 was initially considered to represent a reasonable level of reduction in the data set. The original 7 cluster classification generated 165 NSAs, this number increased to 227 when 9 clusters were defined.

In the interests of direct comparability new areas of similar population size should be achieved, but since EDs themselves have widely differing populations (from 18 to 1,642 residents) it is not surprising that the NSAs varied between 41 and 34,868 people. Unfortunately, this difference is so large that to attempt to reduce it by subdividing areas with the highest populations would result in many extra NSAs being formed. Alternatively, aggregation of the smaller NSAs, such as those comprising single EDs is intuitively unsatisfactory and unless done on a large scale would be unlikely to yield a significantly different result.

The stability of the classification has been demonstrated by a combination of diagnostic statistics and cluster descriptions earlier in this chapter. Following several different starting points, both the 7 and 9 cluster classifications consistently produced similar results.

Finally, a 'good' or 'useful' classification should produce a set of clusters with easily distinguishable and meaningful characteristics, while a 'poor' classification will render a pattern that is impossible to interpret. The 9 cluster classification eventually selected had a distinctive set of clusters which corresponded well with empirical knowledge and expectations and performed satisfactorily according to the above criteria.

#### **Step 7. Cluster mapping.**

Cluster 1 was loosely described as 'established private



residential' - a weak label for the least distinctive of the clusters. The distribution of its constituent NSAs (Figure 5.5), scattered as they are throughout the study area, provides no further information.

The map of Cluster 2 ("low status, council" - Figure 5.6) clearly identifies the city's main council estates at New Parks, Braunstone, Beaumont Leys, Saffron Lane, Spinney Hills, Crown Hills and Thurnby Lodge. All of the major NSAs in this cluster lie within the city boundary.

The map of Cluster 3 ("upwardly transient" - Figure 5.7) shows a marked spatial concentration around the city centre. Many of the NSAs are small, comprising the densely populated EDs which are often found in an inner city area. The range of variables associated with this cluster suggests a classic image of shared accommodation, with young, single, working people moving temporarily into 'bedsitterland' at the start of their working careers. The map of this cluster adds the important dimension of centrality to this description.

Cluster 4 ("low status, ethnic quarter" - Figure 5.8) has an equally distinctive distribution with a clear 'wedge' appearing to the north east of the city centre. As almost all of the 107 EDs that comprise this cluster are contiguous there are very few NSAs. This indicates the degree to which the Asian community in Leicester is spatially concentrated.

In the light of the distribution of NSAs in Cluster 5 ("high social status, low density" - Figure 5.9) it is possible that this cluster should be re-labelled 'suburban'. The majority of NSAs lie outside the city boundary, almost encircling the study area. A 'life cycle' explanation may be offered for this distribution, it shows the results of similar decisions by families to move outwards to newer and more spacious accommodation than would be



Figure 5.5

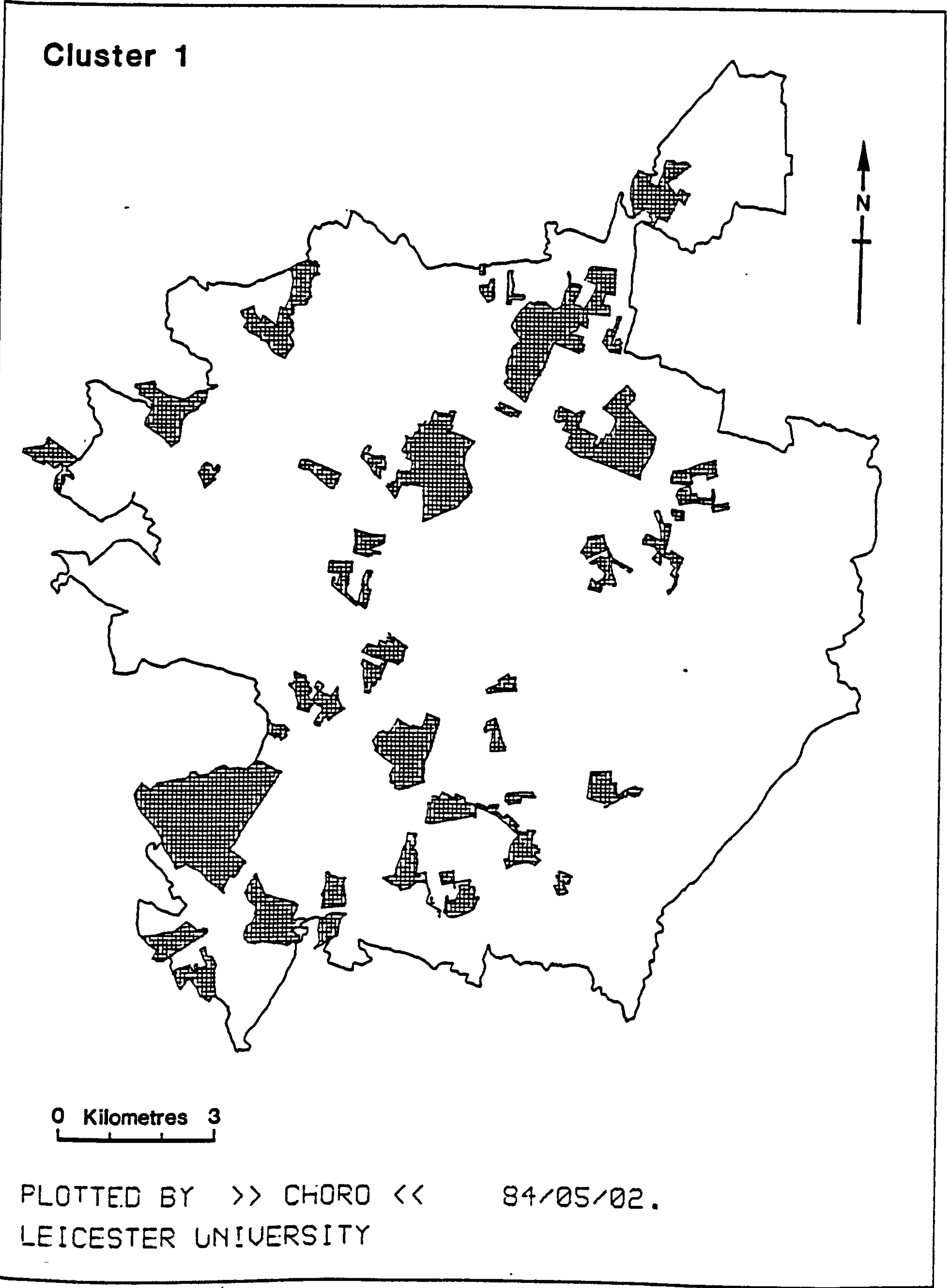


Figure 5.6

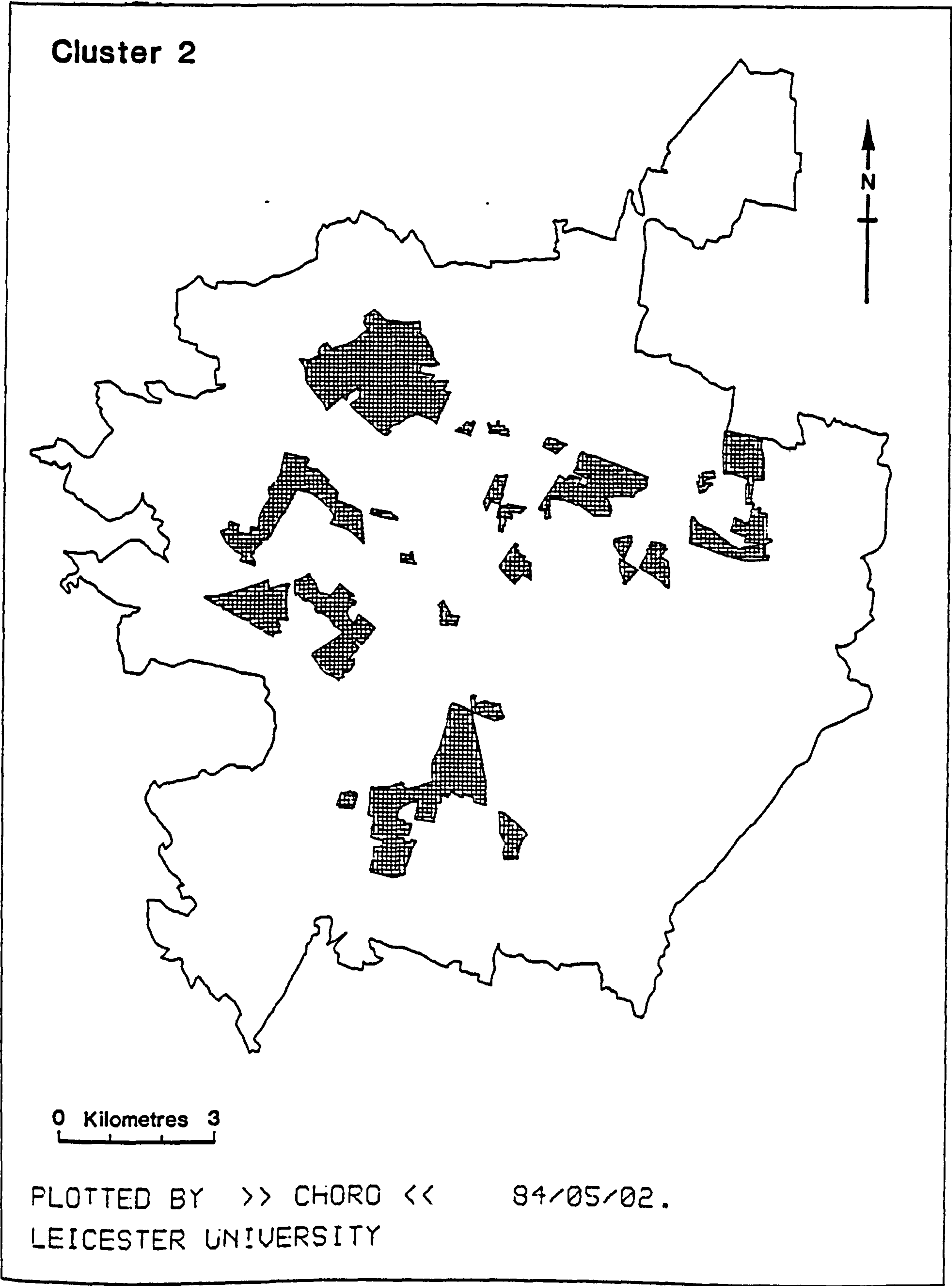


Figure 5.7

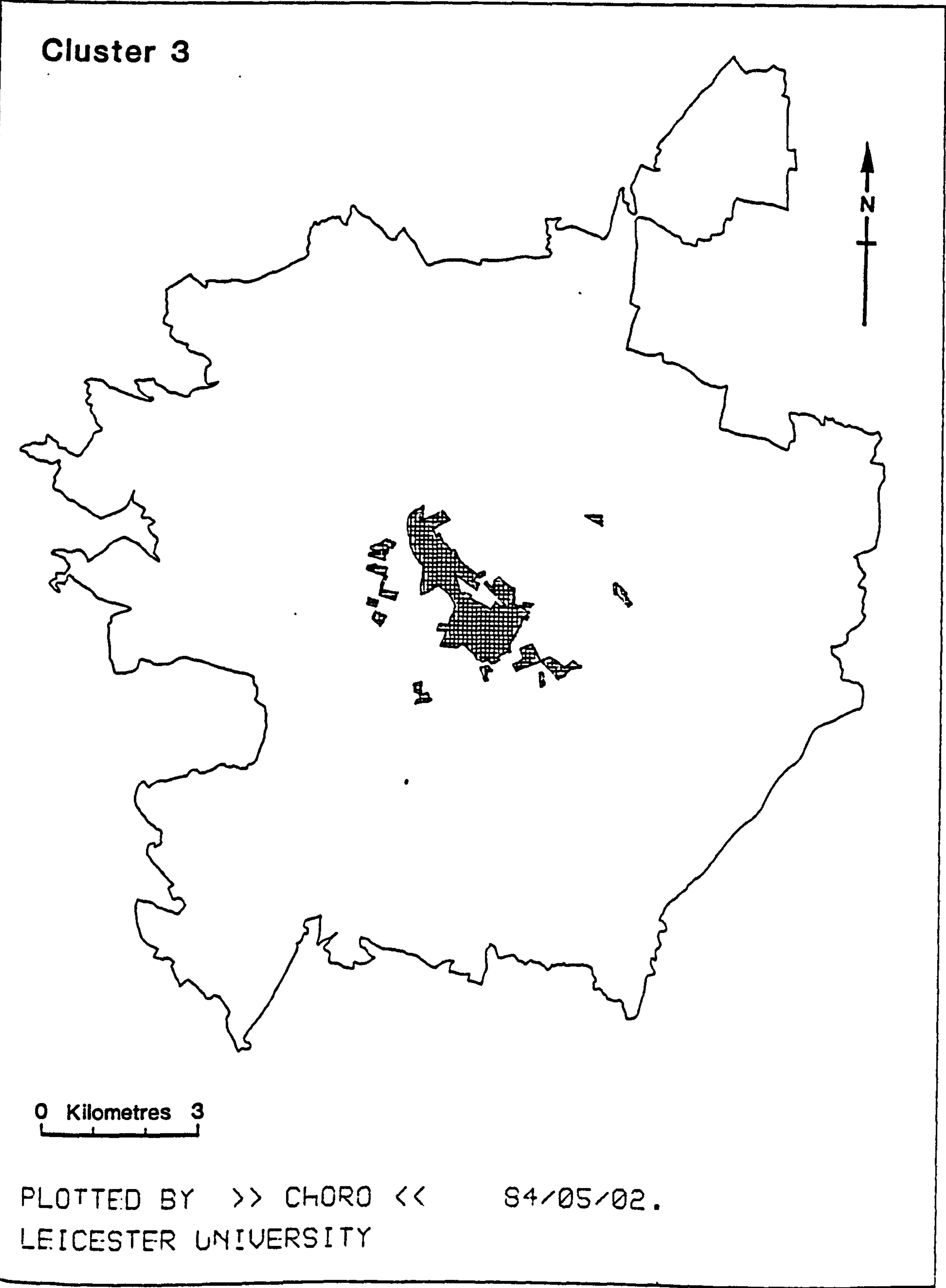




Figure 5.8

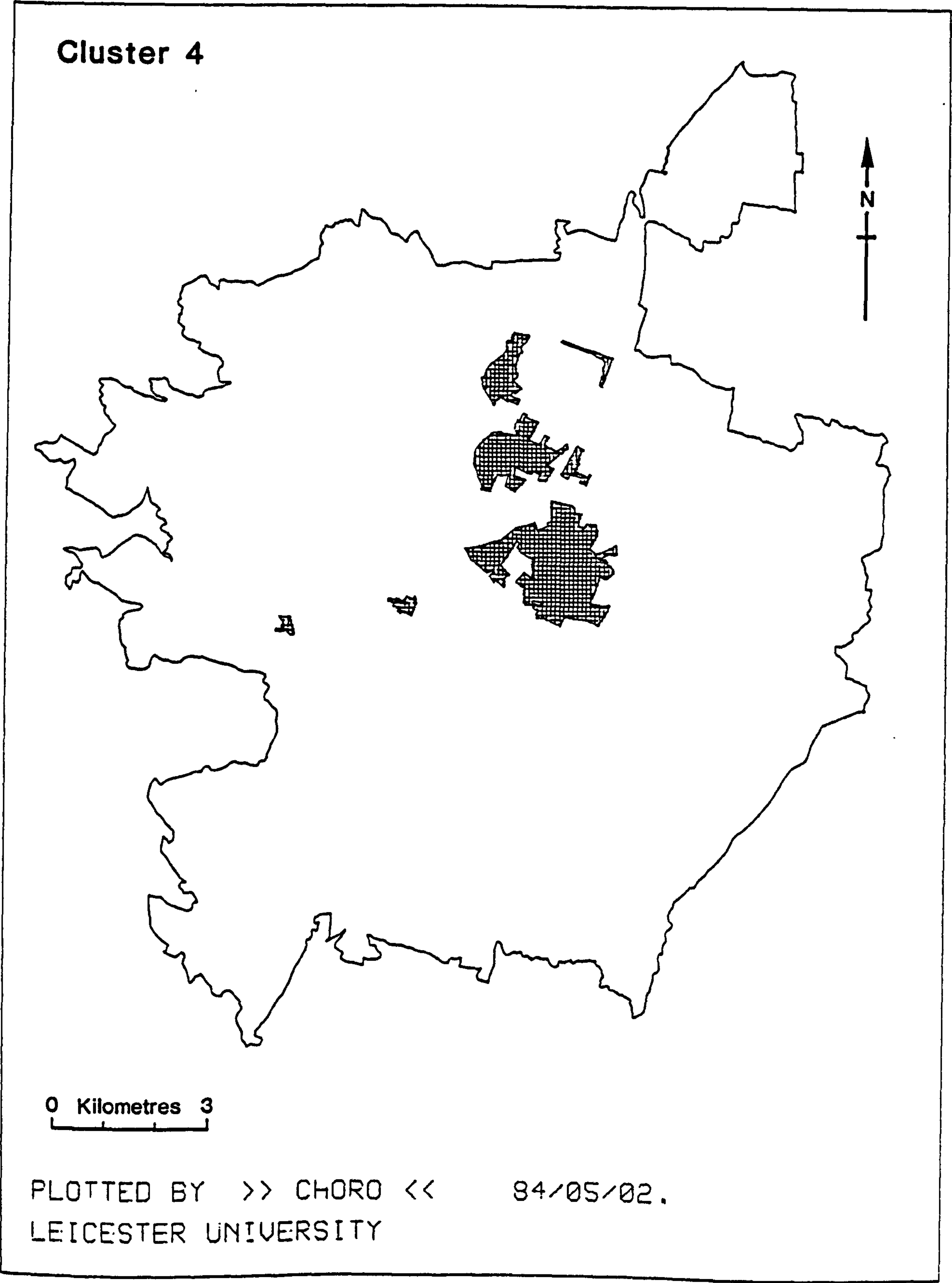
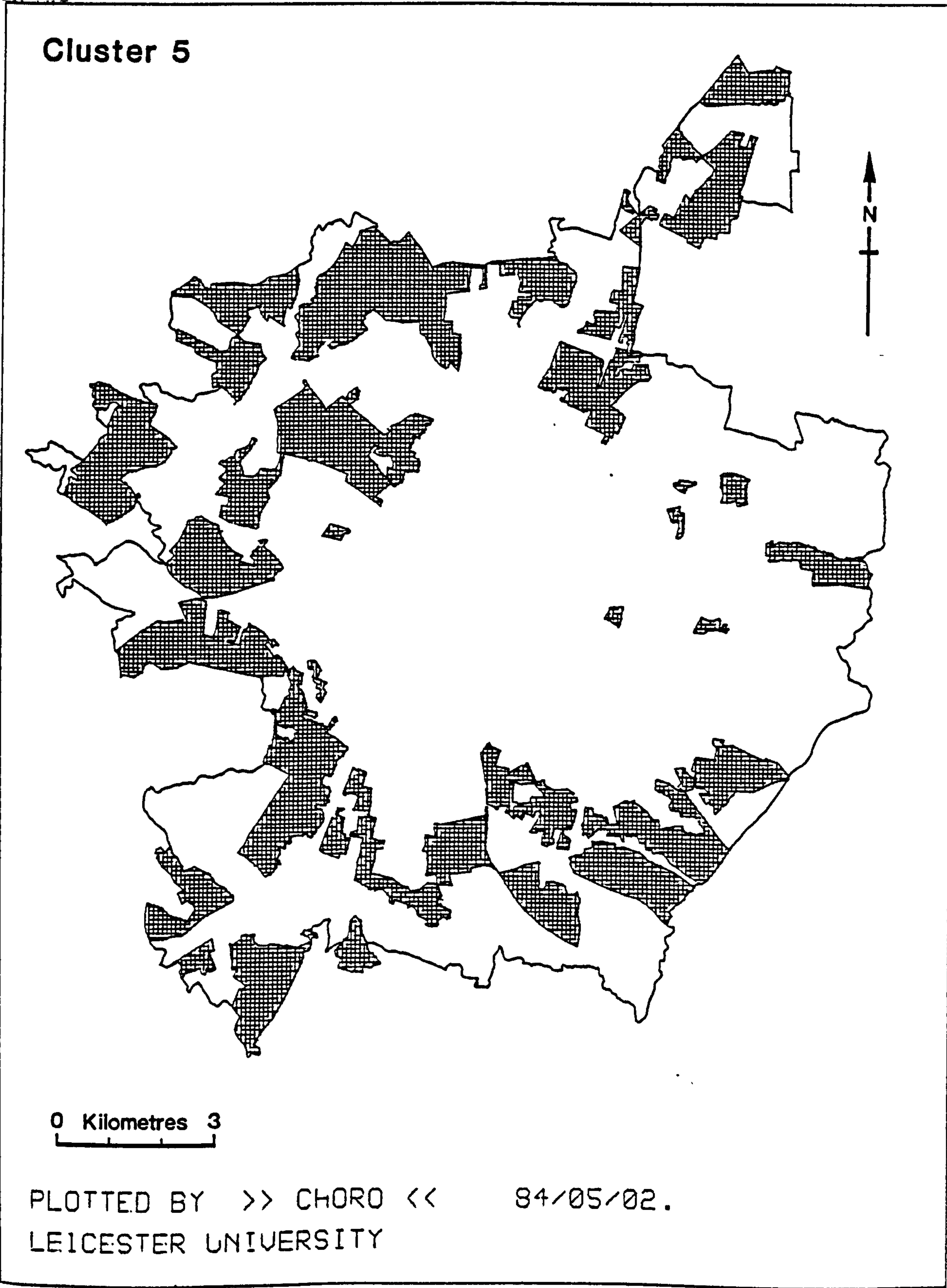


Figure 5.9



found nearer the city centre.

Cluster 6 (Figure 5.10) has been described as "socially dependant, council", the most appropriate label for a cluster which produced many diagnostic variables. Although the NSAs within this cluster are distributed across the city and show no immediately obvious pattern, they may be explained by the recent history of council property development. Between 1964 and 1982 almost half the new council properties have been one-bedroom dwellings or bedsitters, most of these have been purpose-built flats. Two main types of site have been used by the city council. Firstly there have been 'infill' sites, predominantly occurring as a result of clearance and redevelopment near the city centre. These have largely been used for high-density, low-rise flats. Secondly, existing council-owned land has been developed, for example vacant sites on the major council estates. The latter have been used for new bungalow developments for elderly people. Both types of site can be seen in Cluster 6.

NSAs in Cluster 7 ("high socio-economic status" - Figure 5.11) are found mainly to the south east and east of the city, traditionally the more affluent parts of Leicester. Elsewhere, smaller NSAs are to be found at East Goscote, Birstall, Anstey, Glenfield, Kirby Muxloe and along the line of the river to the south west of the study area.

The NSAs in Cluster 8 ("elderly population" - Figure 5.12) are generally small and widely scattered, suggesting only that location is not an important dimension of this cluster. It is unlikely that a cluster dominated by the elderly would have anything other than a fragmented distribution since the circumstances and residential decisions resulting in the spatial pattern would vary so greatly.



Figure 5.10

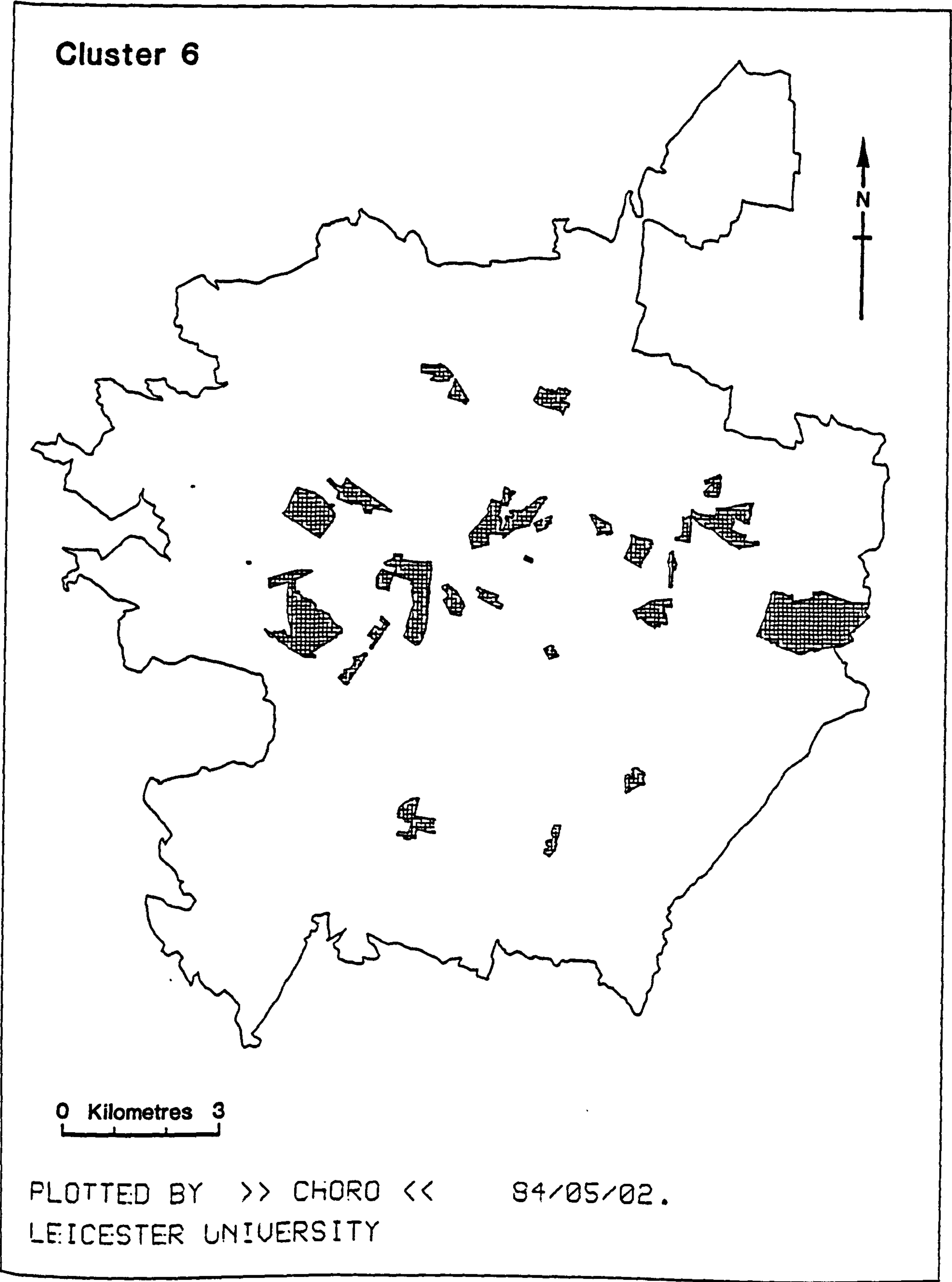


Figure 5.11

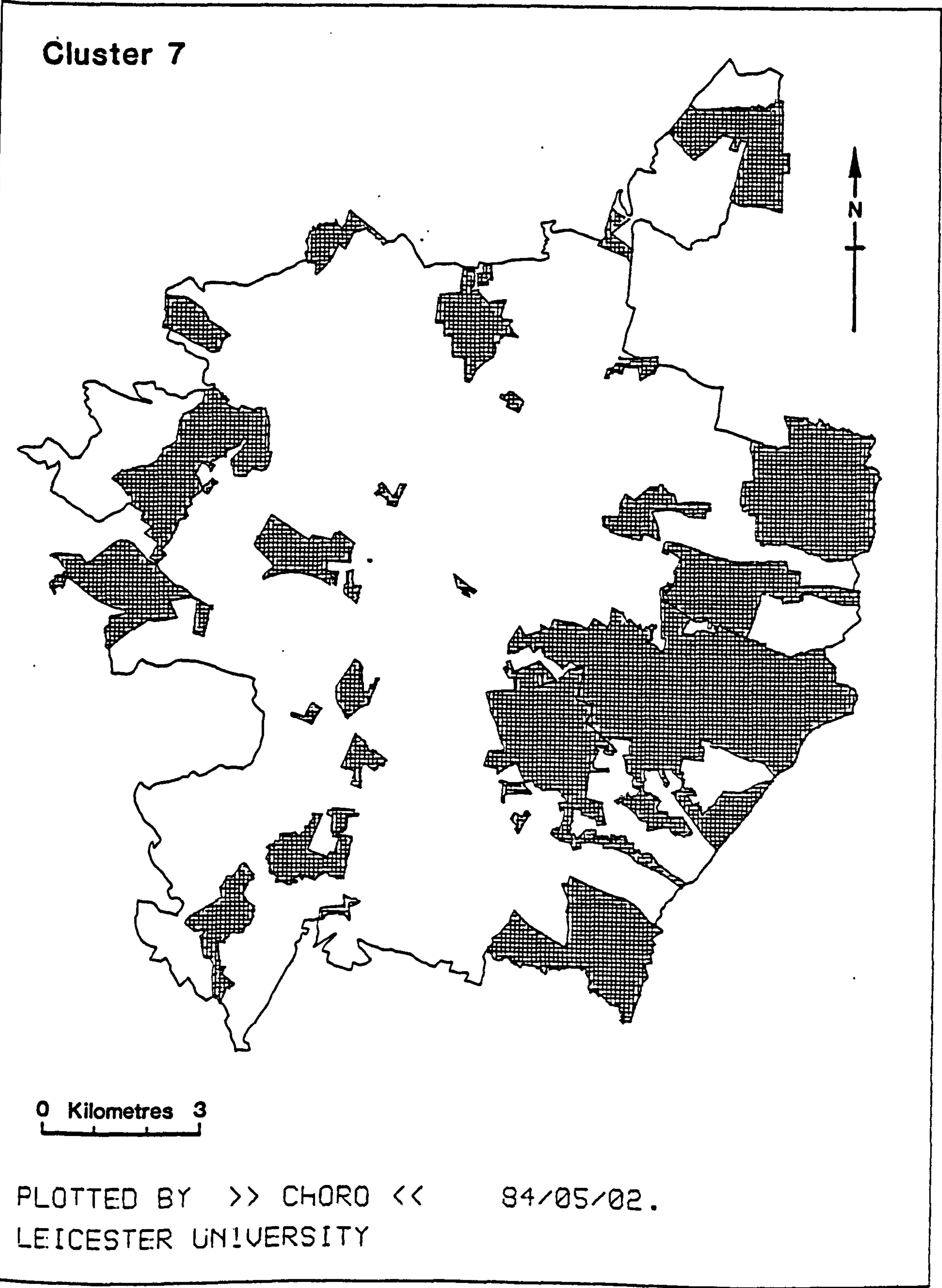
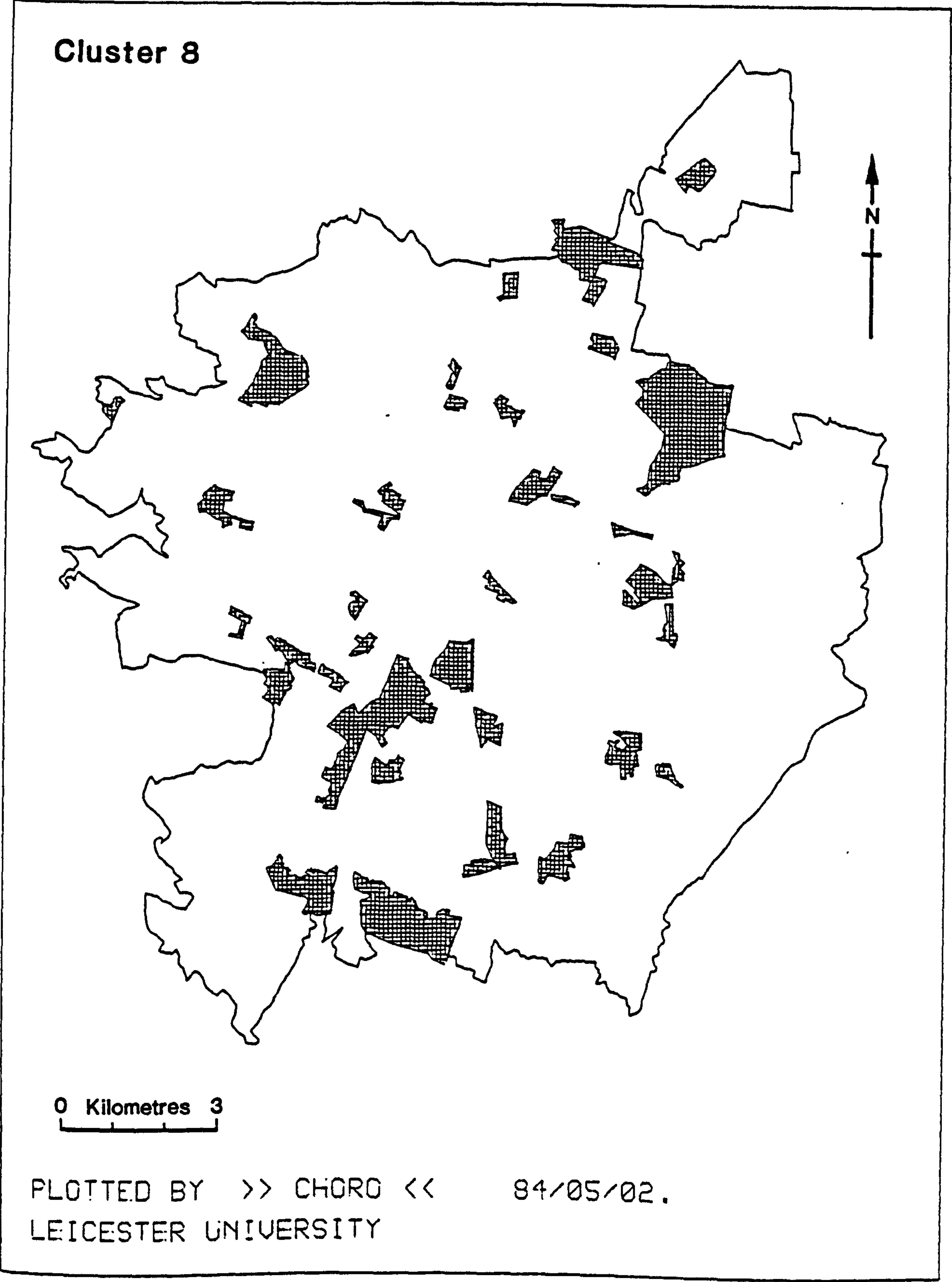


Figure 5.12





Cluster 9 ("private renting" - Figure 5.13) pinpoints those parts of nineteenth century Leicester not favoured by Asian families who opt for owner-occupation.

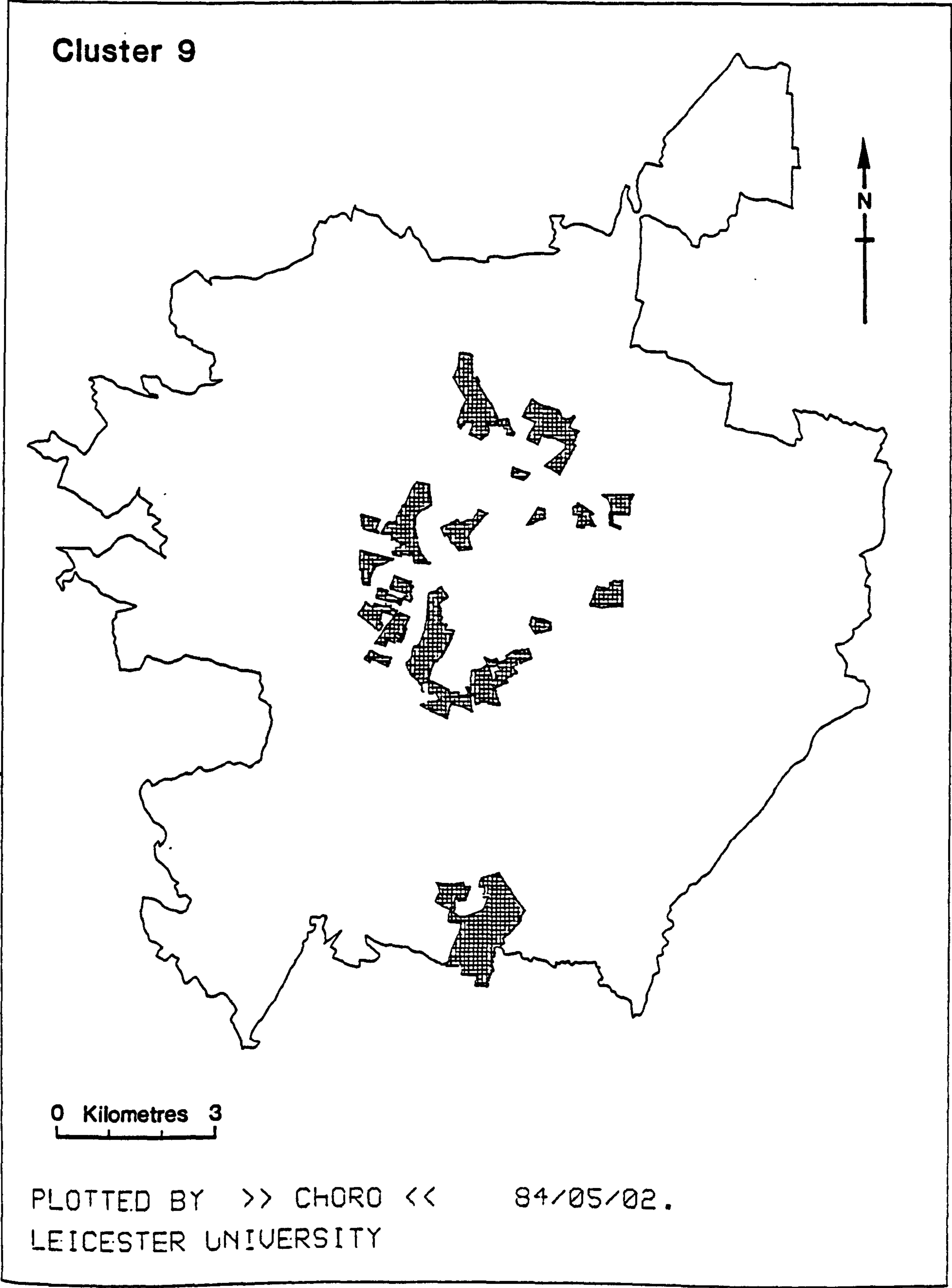
Although no measure of location was included in the cluster analysis, it is clear that most clusters have a marked spatial component. The cluster maps are important tools in supporting and clarifying the earlier descriptions.

## 5.5 Conclusion.

This chapter's detailed discussion of the procedure and results of the classification of EDs in Greater Leicester provides a foundation for the remainder of this thesis. Nine clusters have been identified, each comprising a set of EDs with distinctive social, economic and demographic characteristics. They take the following labels: "established private residential", "low status council", "upwardly transient", "low status ethnic quarter", "high social status, low density", "socially dependant, council", "high socio-economic status", "elderly population" and "private renting". EDs in each of these clusters have been combined on the basis of spatial contiguity and the resulting 'new small areas' (NSAs) have been mapped, showing that in most cases clusters have a distinct spatial dimension.

In the chapters that follow, new census data will be extracted for the NSAs and these results will be compared with data from the Perinatal Mortality Survey. In combination they will provide a comprehensive picture of the urban environment into which the baby is born.

Figure 5.13



## Chapter 6. A census map analysis of Greater Leicester

### 6.1 Introduction

The classification of census EDs into new small areas was based on a set of 34 social, demographic and economic variables from the census. In this chapter each of these variables will be considered in greater detail in order both to describe the study area in general terms and to highlight potential areas of disadvantage. The basis of this discussion is a set of maps displaying chi-square values for census variables. The decision to use this form of data transformation is justified in Appendix 1. The areal analysis will provide a basis for comparison with the perinatal data in the subsequent chapter. First, the recent history of the city of Leicester is discussed.

### 6.2 History of Leicester : Summary

Although parts of modern Leicester can be traced as far back as Roman times most authors agree (Evans, 1970; Simmons, 1974; Lewis and Davies, 1974) that today's city structure is a nineteenth century product. Between 1801 and 1901 the population of the city grew rapidly from 17,000 to 211,000 persons (Evans, 1972) with consequent impact on social, economic and physical structure.

Reasons for this unbridled population growth relate primarily to Leicester's developing role as a manufacturing centre. Firstly hosiery, and subsequently the footwear and engineering industries were responsible for both attracting migrants from rural areas, and later supporting a high rate of natural increase.

Geographically, the development of the city was constrained by five major principles:

"site constraints; the availability of land and land ownership; the extent of municipal control; the pre-empting of the most desirable land by the wealthier citizens of the city; and the



perpetuation rather than the replacement of distinctive social areas within various parts of the city" (Lewis and Davies, 1974, p. 195).

The major site constraint was the River Soar and its floodplain. Whilst population pressure had been slight it was possible to avoid building on those areas subject to periodic flooding, but following massive population growth they became covered with the cheapest working class homes.

The availability of land and land ownership were particularly important to Leicester's growth in that, on the one hand, building was not restricted by surrounding town fields (as was the case in Coventry and Nottingham), but on the other hand, some powerful landowners were able to withstand pressure from developers and hence limit growth, for a while, to certain directions. In particular, to the north the Abbey and St. Margaret's pasture provided a block to the spread of the town; to the west (across the river) the estates of Westcotes and Dannel's Hall resisted development until 1886 and 1861 respectively; while to the south, although the land had been enclosed for some time, it was mostly held by the corporation and reserved for public purposes (Evans, 1972). In this area the cemetery, cattle market, gas works, university and the old racecourse are now to be found. Only to the east was land readily available. The east fields had been enclosed in 1764, but divided into small lots subsequently owned by individuals who were grateful to sell to developers. As a result the entire area became swamped by a mass of high density housing, and St. Margaret's parish (in which it all lay) by 1851 contained two-thirds of the town's population (Evans, 1972).

A significant trend during the late nineteenth century was the outward migration of the wealthier people. Following the replacement of the domestic hosiery industry by the factory system, space for new industrial building was required and often

was supplied by wealthy manufacturers leaving their homes in the town centre for new properties at the periphery. (Evans, 1972). Rising standards of house building accompanied rising incomes and there was an increased awareness of housing as a symbol of social and economic status (Pritchard, 1976). A distinct sector of high class housing was perpetuated outwards towards the south east.

In response to the late nineteenth century expansion Leicester's borough boundaries were in 1892 extended to include Aylestone, Knighton, Belgrave and large parts of Humberstone and Evington parishes. This effectively trebled the municipal area and made Leicester one of the fifteen largest towns in England.

The twentieth century has seen both the continuation of processes started in the nineteenth century (for example improvements in transportation, the increased diversification of industry and suburbanisation of the population) as well as the initiation of new schemes. Of particular significance have been the development of municipal housing estates chiefly in peripheral locations; the growth in service sector employment; the private development of areas of high status housing, particularly in village locations such as Oadby, Wigston, Blaby and Birstall; and most recently, the rejuvenation of some inner city areas as a result of occupation by primarily New Commonwealth immigrants.

Despite subsequent boundary changes, most recently in 1966, the population of Leicester city has shown an overall decrease during the post-war period (see Table 6.1). This has been due not to a natural decrease, but rather to net emigration, to suburban areas both within and outside the city boundary. Both public and private housing schemes have contributed to the general out-movement. Within the city, municipal housing developments have been most significant - since 1946 some 16,591 dwellings have been

**Table 6.1                      Population change in Leicester City**  
**1951 to 1981**

Year	Population	Change
1951	285181	
1961	273470	-11711
1971	284210	+10740
1977	281000	-3210
1981	280324	-676

**Source: Leicester City Council, 1984**



**Table 6.2 Greater Leicester : Facts and Figures**

Area 165.83 km²

(i) Including all enumeration districts: (City totals in parentheses)

Total Population (all residents) 428086 (276245)

Total Households (with residents) 151985 (92732)

**Age structure:**

0-4 yrs.	27738	(19033)
5-9	28715	(18191)
10-14	35424	(22260)
15-19	37627	(24352)
20-24	35685	(25627)
25-29	30808	(21078)
30-34	30836	(18458)
35-39	24882	(14035)
40-44	23580	(13762)
45-49	23520	(13743)
50-54	24314	(14976)
55-59	24849	(16059)
60-64	20557	(13623)
65-69	19384	(12937)
70-74	16669	(11492)
75-79	12316	(8645)
80-84	6969	(4947)
85 and over	4213	(3027)

(ii) Excluding special enumeration districts:

Total population (all residents) 425872

Total households (with residents) 151712

**Households:**

Single person	32860
Lone parent	3591
Owner-occupied	92715
Council rented	41326
Privately rented	12245
Lacking use of amenities	10332
With more than 1.5 persons per room	1313
Without a car	63570

**Economically active population**

Total aged 16+	211673
Aged 16+ seeking employment	20185
Aged 16 to 24 out of employment	7567

**Socio-economic status (10% sample)**

Professional and managerial	2282
Non-manual	2436
Skilled manual	4542
Semi-skilled	2358
Unskilled	602

Persons born in the New Commonwealth or Pakistan 44844

One year migrants 39547

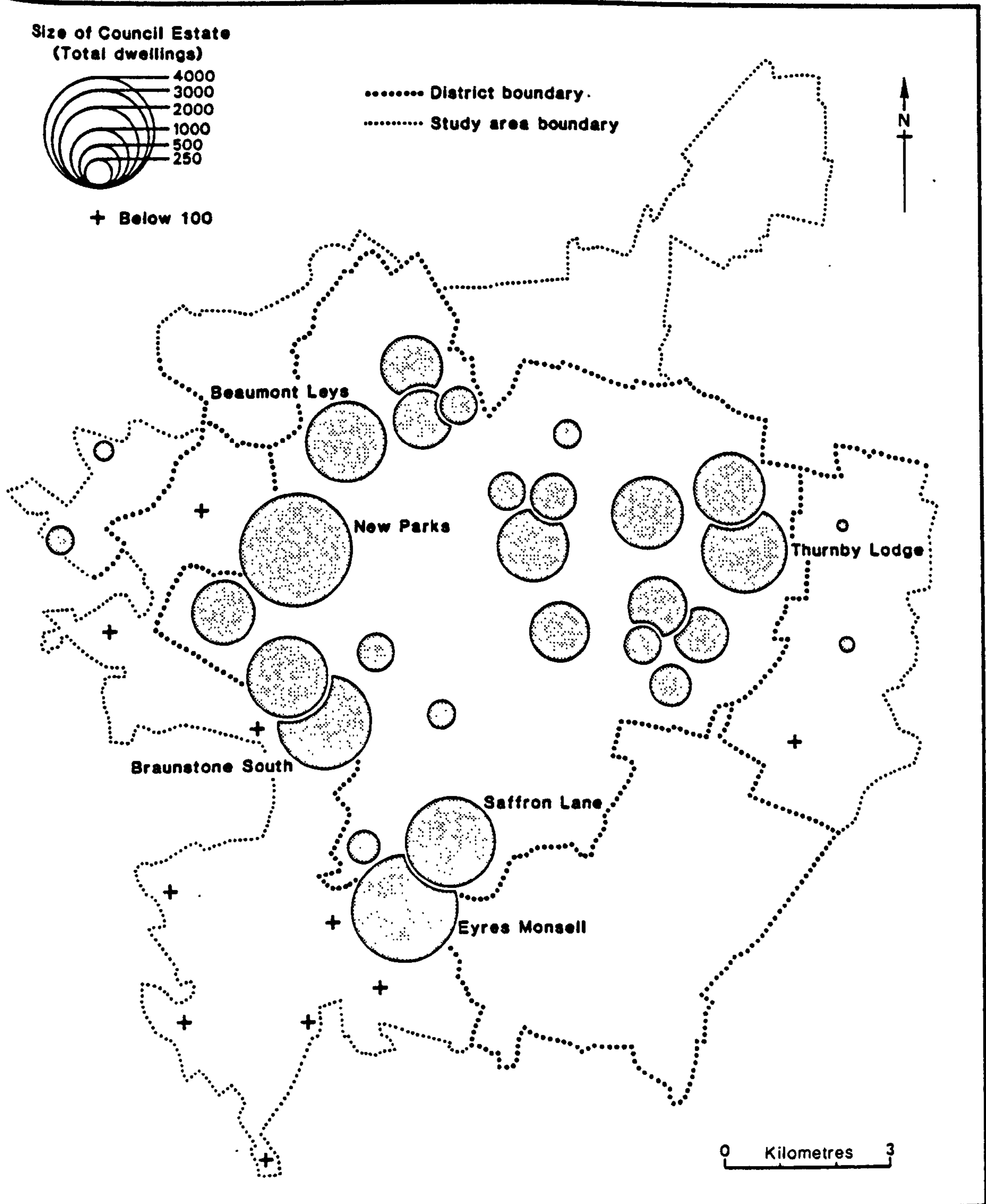
(Source: 1981 Census)

added to the council housing stock. The largest estates of this period are to be found on the periphery of the city at Eyres Monse11, New Parks, Thurnby Lodge and Beaumont Leys (see Figure 6.1). Private developers, on the other hand, have concentrated their activities on high status housing just outside the city boundary. The 'flight to the countryside' (Lewis, 1972) has resulted in the engulfment of villages such as Oadby, Wigston, Thurnby and Birstall by the rapidly expanding urban area (Strachan, 1972).

Although the direction of population movement has been in the main away from the city centre (Strachan, 1972), there has been some movement towards the inner city. Two explanations are offered for this. Firstly, the city council have had both a direct and an indirect influence. Directly, they have been responsible for redevelopment of the inner city and the creation of new housing estates there such as St. Peter's, St. Mark's and St. Matthew's; indirectly they have provided incentives to inner area residents through grants for home improvements. Its 'housing renewal strategy' has been designed to improve the quality of both the housing stock and the environment, and to alleviate the otherwise resulting social stress in the inner area.

The second major force in the preservation of the inner city has been the recently established Asian community. It has been estimated (Leicester City Council, 1984) that in 1983 the total Asian population in Leicester was 63,186; comprising some 42,222 Gujaratis, 11,455 Punjabis and significant numbers of Kutchi-, Bengali-, Hindi- and Urdu-speaking people. Most of these people are recent (post-war) immigrants to the city. Originally attracted by the post-war boom in employment opportunities, and subsequently arriving following expulsion from Uganda, the ethnic community is now firmly established in the Highfields and Belgrave areas.

Figure 6.1 Greater Leicester - Council Housing





Economically, Leicester has traditionally benefitted from a diverse industrial base which is concentrated within the city along the river and canal on a north/south axis through the city centre. Although there has been a decline in the older industries (textiles and footwear), for a while the expansion of light engineering compensated for this loss in employment (Evans, 1972). Recently, and in common with most other British cities, the service sector has shown the greatest expansion.

Improvements in transportation have included the replanning of roads, increasing accessibility both within the city and between city and county (Ellis, 1976). The city continues to dominate the county economically, socially and numerically (Strachan, 1972) and this position is unlikely to change. It retains its importance as a self-contained market centre (Simmons, 1972).

### **6.3 Univariate descriptions**

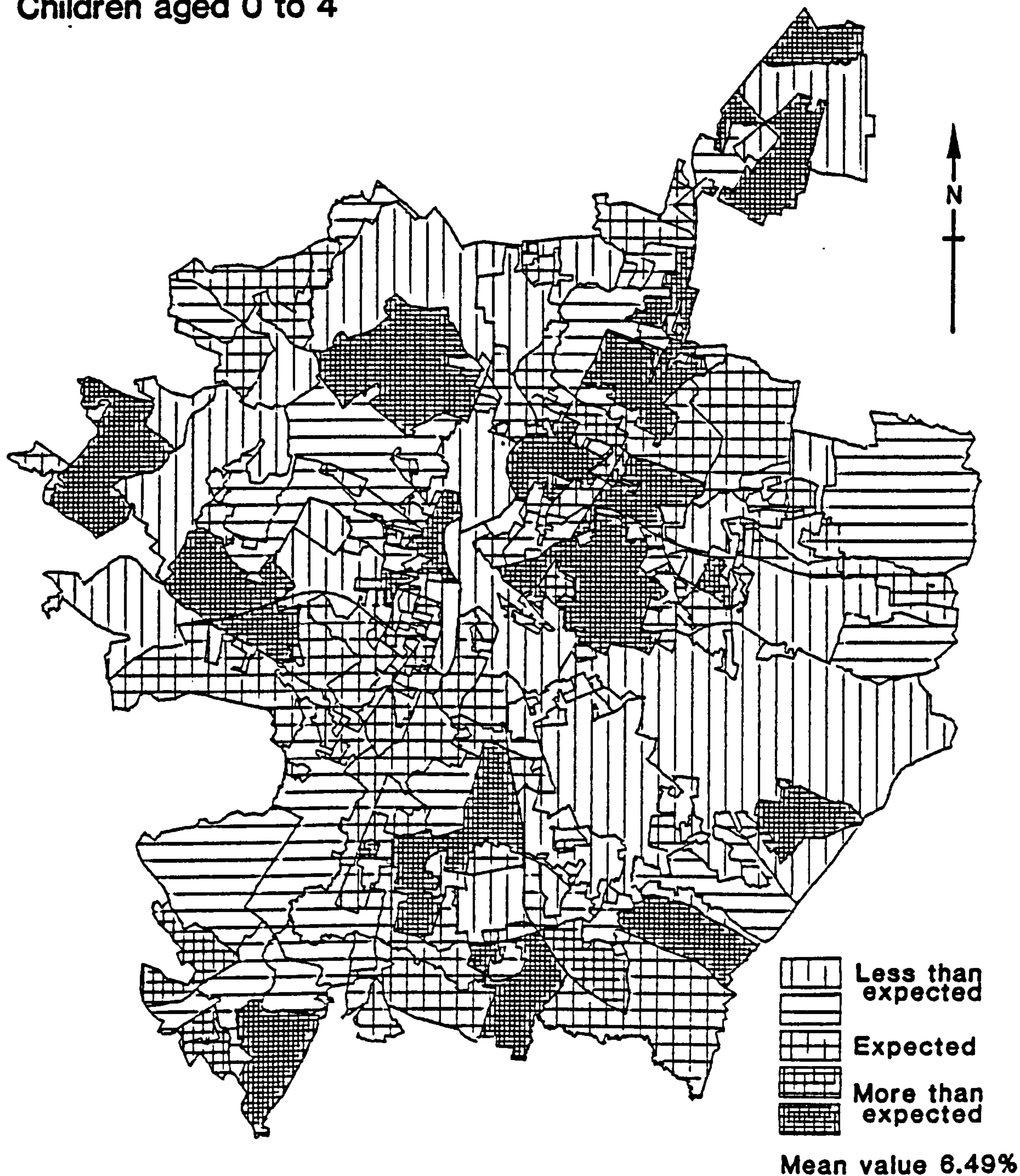
#### **6.3.1 Demographic characteristics**

Seven demographic variables were included in the data set, each referring to an age group: 0 to 4 years, 5 to 15, 16 to 24, 25 to 44, 45 to retirement, retirement to 74 and 75 or more. Figures 6.2 to 6.6, 6.8 and 6.9 show the distribution over Greater Leicester of each of these age groups.

**Children aged 0 to 4 (Figure 6.2).**

A wedge of dense shading leading from the city centre in a north-easterly direction and a scattering of densely shaded areas in the outer suburbs dominate this map. The presence of young children in a family represents a new stage in the life cycle and new requirements of space and accessibility. In Britain these requirements may be met by either public or private housing provision. In Greater Leicester the peripheral council estates at

Children aged 0 to 4



0 Kilometres 3

PLOTTED BY >> CHORO << 94/04/05.  
LEICESTER UNIVERSITY



Mowmacre Hill, Braunstone, Eyres Monsell and South Wigston all have more children in this age group than 'expected'. Given that the allocation of council houses favours families with children this pattern is not surprising. The other suburban areas highlighted on this map are predominantly zones of new privately built housing. At Groby, Cosby, Oadby, Syston and East Goscote there has been considerable development recently of relatively cheap and low density accommodation. In South Wigston there is in addition some privately rented property which attracts young families unable to either buy their own home or qualify for a council house.

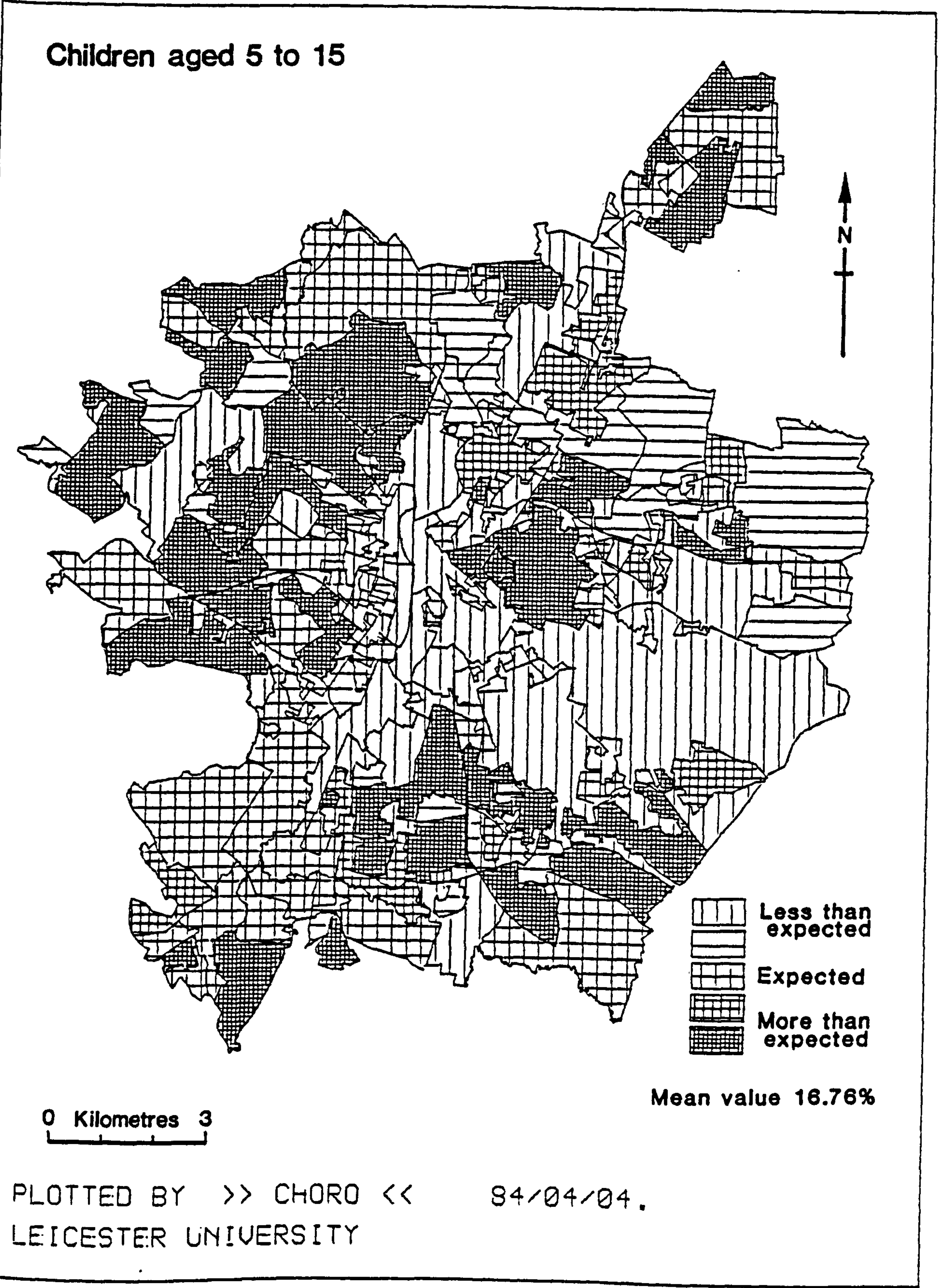
The high density sector extending to the north-east is explained by the presence there of a large ethnic minority community. As explained previously, the post-war period has seen the establishment and rapid expansion of the city's Asian population firstly in the Highfields area and subsequently in Belgrave. The existence in these areas of a comparatively high proportion of young children reflects both the age structure of the immigrant population and their socio-cultural preference for the larger family size. This creates a direct contrast to the expected elderly age structure conventionally found in the central areas of other British cities.

#### Children aged 5 to 15 (Figure 6.3).

The presence of older children within a household is indicative of a more established family structure. Spatially, such households are likely to be found in the older suburban areas. In Greater Leicester the inter-war council estates at Braunstone, Northfields and Saffron Lane, and the earlier post-war estates at New Parks and Mowmacre Hill have high proportions of children in this age group, so too do some of the older private developments, especially those centred on existing villages such as Oadby,



Figure 6.3



Wigston, Blaby and Thurmaston. There are strong similarities between Figures 6.3 and 6.2 showing younger children.

**Persons aged 16 to 24 (Figure 6.4).**

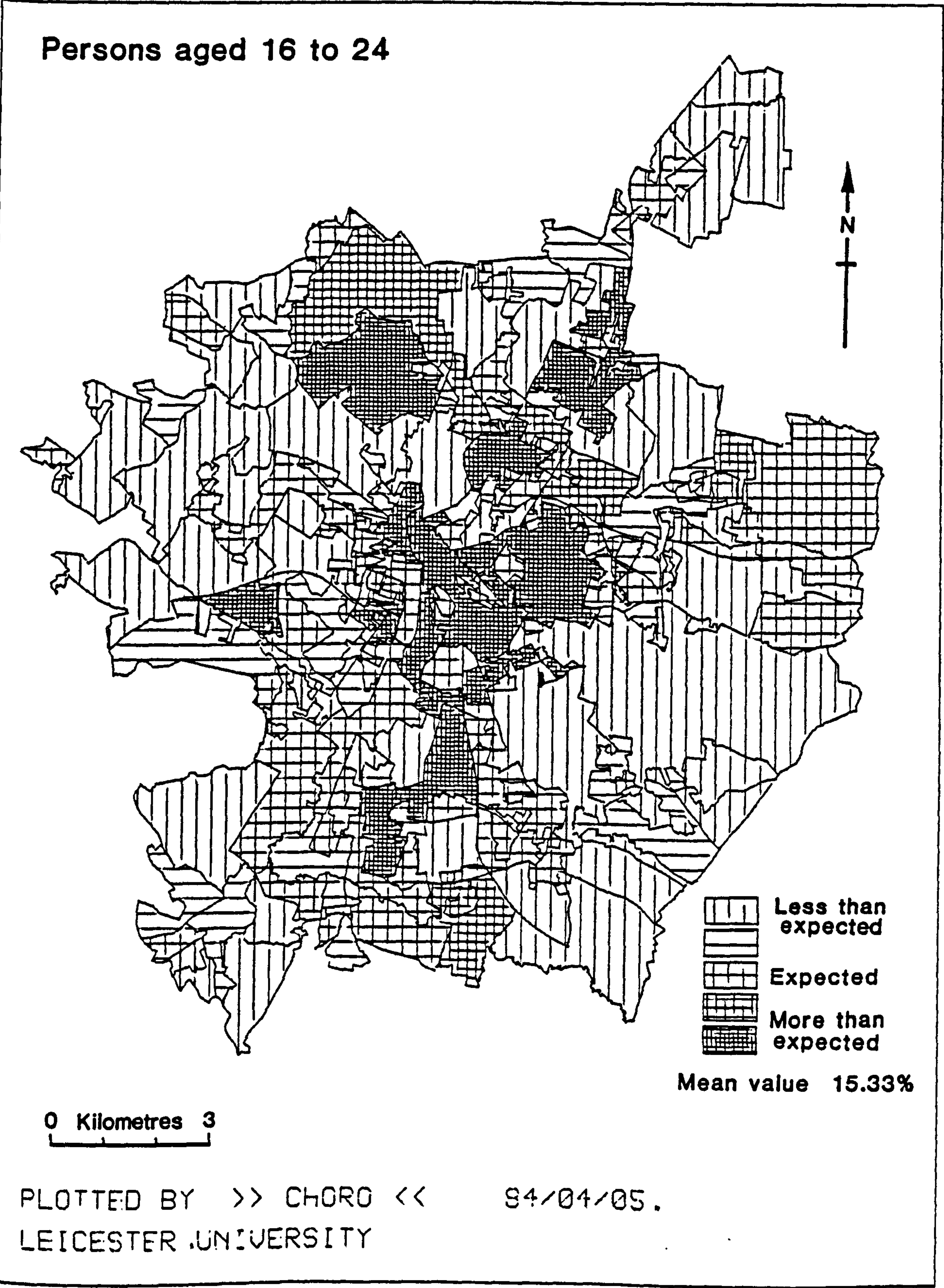
The predominant concentration of young adults is around the city centre and the Spinney Hill and North Evington areas. The inner city is particularly important as a residential location for young people leaving the family home for perhaps the first time and seeking cheap and shared accommodation. As with the younger age groups, life cycle stage plays a major role in residential decision-making. The availability of bedsits, converted flats and purpose-built council flats all make the central area a desirable location for this age group.

A second concentration occurs along the Highfields/Rushey Mead sector, as before reflecting the age structure of the Asian population. Highfields was the first part of Leicester to be settled by the Asian community and the high proportion of young adults there in 1981 arises from the subsequent maturity of children born in the early years. Due to pressure on housing resources there has been sectoral out-movement of the ethnic community, first to Belgrave and more recently to Rushey Mead. This movement has involved a significant number of young people seeking to establish their own households, hence the dominance of this age group in Rushey Mead. Also found in Rushey Mead is an established white community of middle income families which completes the picture.

The remaining parts of Greater Leicester having above average proportions of young adults are all older council estates, namely North Braunstone, Saffron Lane and Mowmacre Hill. In these areas the young people concerned have probably not yet left the family home, as with the Asian population today's demographic



Figure 6.4





pattern reflects earlier housing decisions.

#### **Persons aged 25 to 44 (Figure 6.5).**

Numerically the 25 to 44 age group comprises the largest demographic category, having some 26% of the total population. Two sub-groups may be identified: those with and those without family commitments. With regard to the former much larger group, favoured residential locations will probably include the lower density suburban areas, while the latter may choose the advantages of a central location. Both preferences are represented in Greater Leicester.

As shown in Figure 6.5, the greatest concentrations of this age group are found in the newer private estates at the periphery of the study area. They almost encircle the city. It is notable that nearly all of the densely shaded areas are located beyond the city boundary. This indicates both the importance of suburbanisation as a process in Greater Leicester, and also the paucity of the fixed boundary definition for a demographic study.

Within the city, few areas have above average proportions of 25 to 44 year olds, in the suburbs only the newest developments near Beaumont Leys, Western Park and Rushey Mead can be identified. The city centre, on the other hand, has a relatively high concentration of this age group although this must be seen within the context of a relatively small total population. It is suggested that in the central area will be found predominantly single people and those couples who have yet to start a family. The availability in this area of a large number of Victorian terraced properties for both rent and first time purchase provide an attractive inlet to the housing market.

#### **Persons aged 45 to retirement (Figure 6.6).**

With certain exceptions, this map displays a pattern that is the inverse of Figure 6.2 showing the distribution of young

Figure 6.5

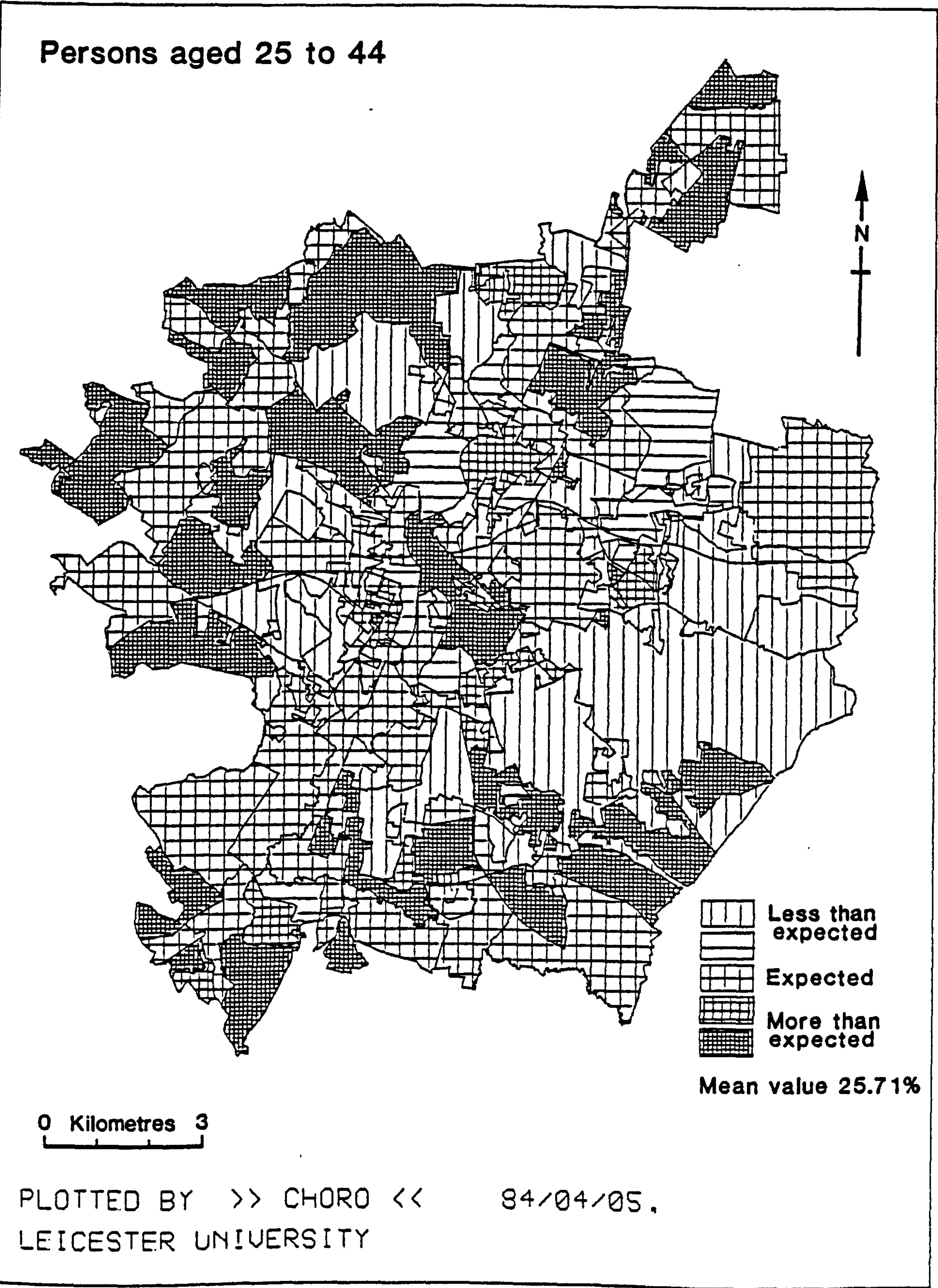
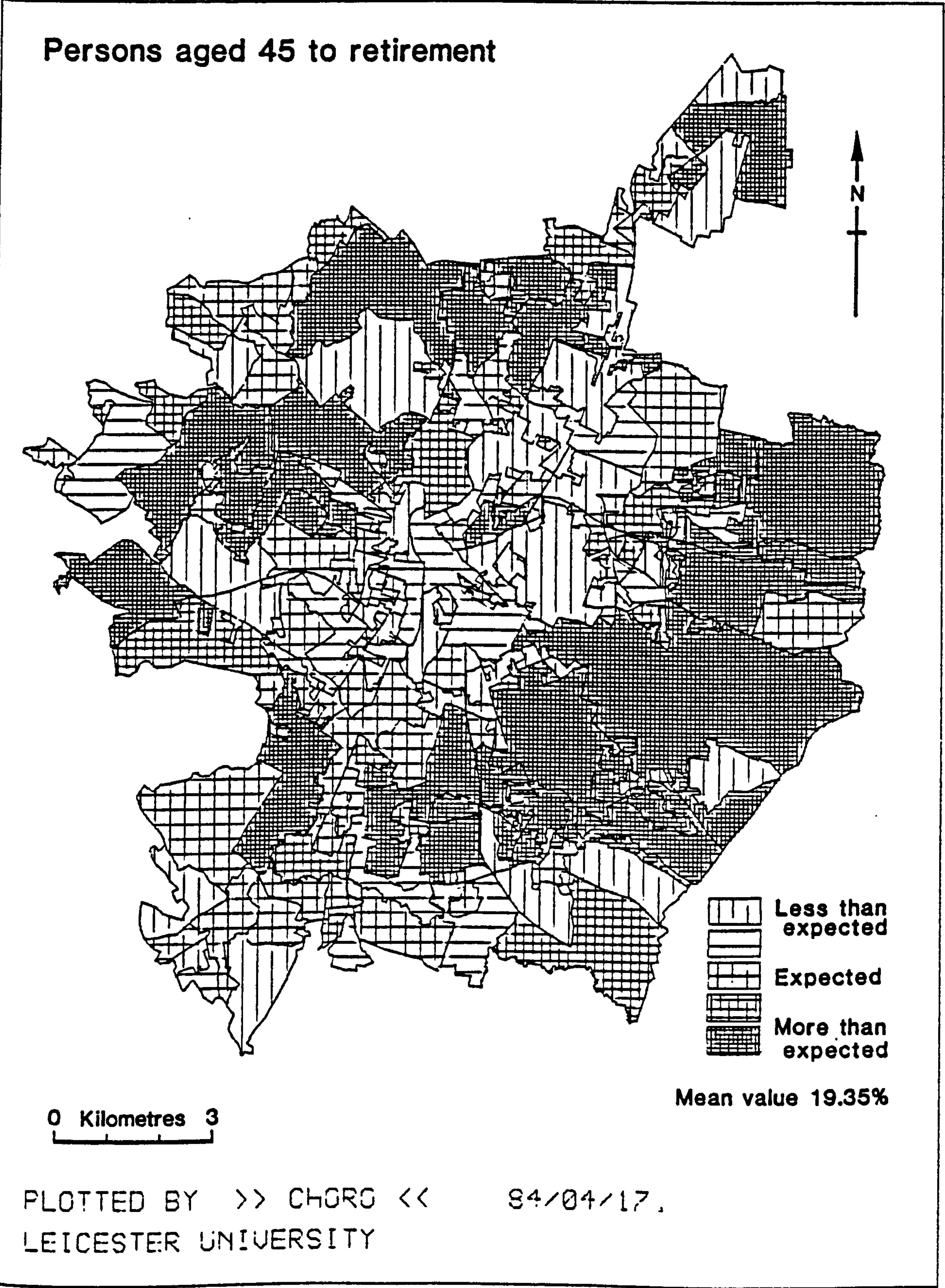




Figure 6.6





children. This of course is to be expected if they are considered in the context of the normal life cycle pattern.

Across the study area as a whole, some 19% of the population fall into this age group, with minimum and maximum figures of 1.9% and 43.9% being attained (see Table 6.3 and Figure 6.7). Significantly, many of the inner suburban zones have high percentages of middle aged people, particularly around the well established commuter villages of the 1950s and 1960s. It is suggested that the present demographic pattern originates in the search behaviour of those people seeking accommodation some twenty to forty years ago. Like today's young adults they probably acquired a residence to either buy or rent on the periphery of the built-up area, but following subsequent expansion of the city they found themselves enveloped by modern development, eventually achieving 'inner suburban' status. Areas of private housing such as Glenfield, Western Park, Knighton, Stoneygate and Evington fall into this category, as do some of the immediate post-war council estates such as New Parks and Eyres Monseil. Once again, council housing allocation policy influences the demographic structure of the city.

For the first time, the eastern suburbs, including Thurnby, Stoughton and Oadby, are prominently shaded on a map. It is suggested that this is directly related to the quality of accommodation and the social status of these areas. Historically there has always been a high status sector leading towards the south east of the city of Leicester (see Section 6.2), this is currently perpetuated by the mechanism of house prices. Thus only those people with significant financial resources would be both willing and able to live in these areas, such people most probably being at least of middle age.

Table 6.3      New small areas with extreme values for census variables

Variable	Minimum value	New small area(NSA)	Maximum value	New small area(NSA)
-----				
Young children	0.0	6E	16.0	2V
Older children	0.6	3B	29.6	4F
Young working	2.4	6E	56.9	3B
Middle working	3.0	6E	40.8	5K
Older working	1.9	3B	43.9	7Y
Young pensioners	0.0	3B	46.4	6E
Older pensioners	0.0	3B	38.2	6E
New Commonwealth born	0.0	7G	51.5	4B
Single person households	2.6	4F	69.9	6E
Large families	0.0	3G	16.8	4F
Lone parent households	0.0	7X	46.3	6W
Mixed households	0.0	6V	33.5	4F
Owner-occupied	0.0	2C	100.0	7CC
Council rented	0.0	7B	100.0	2C
Privately rented	0.0	6P	79.0	3L
Without amenities	0.0	6N	52.2	3L
Overcrowded	0.0	7D	5.8	4F
Not self-contained	0.0	7A	39.5	3L
Purpose-built flats	0.0	7B	100.0	6W
Professional + managerial	0.0	2E	75.0	7G
Non-manual	0.0	2L	61.5	1F
Skilled manual	0.0	7Y	58.8	1R
Semi-skilled manual	0.0	7G	52.9	2D
Unskilled manual	0.0	7A	33.3	6V
Highly educated	0.0	6U	50.0	7Y
Economically active persons	14.4	6E	95.4	3B
Two economically active	3.7	6E	66.0	4C
Unemployed	0.5	7J	37.7	6W
Students	0.0	3B	10.3	7H
Married women at work	23.0	6E	70.0	6U
Young unemployed	0.0	1K	100.0	6V
Self-employed	0.0	3B	21.9	7AA
One year migrants	1.4	1K	50.6	4B
Without a car	2.8	7G	92.6	6E

Note: for some minimum or maximum values several areas could be shown, here only one is included to provide an example.

Figure 6.7 shows the location of NSAs



# Greater Leicester – New Small Areas





**Persons of retirement age and over (Figures 6.8 and 6.9).**

The last two demographic variables are considered together because of the similarity of their distributions. Both have highest densities in the well-established inner suburbs and around some of the older industrial villages, for example Enderby and Blaby.

Following the 1966 sample census, Lewis (1972) first identified a concentric zone around the city centre "characterised by an older population, owner-occupancy, and unfurnished property tenancy, as well as relative under-occupancy" (p.468). It is suggested by Figures 6.8 and 6.9 that, at least as far as the age structure is concerned, such a zone still exists, although it is now disrupted by the presence of the Asian community to the north east.

The redevelopment of the city centre and the provision by the city council of special housing facilities for pensioners further disrupt the natural concentric pattern. For example, the council estates of St. Marks, Gilmorton Avenue, Saffron Lane, Rowlatts Hill, Braunstone, West End, Beaumont Leys and Belgrave all operate warden-assisted schemes for the elderly.

### **6.3.2 Household composition**

Information on the composition of households is frequently used for measuring relative deprivation and for guiding policy making. In particular, households which might make special demands on health and social services are of interest. These households often exhibit spatial concentration as shown on the maps in this section. Five variables are examined here: persons born in the New Commonwealth or Pakistan, single parent families, households with four or more children, single person households, and lastly, households with three or more adults and at least one

Figure 6.8

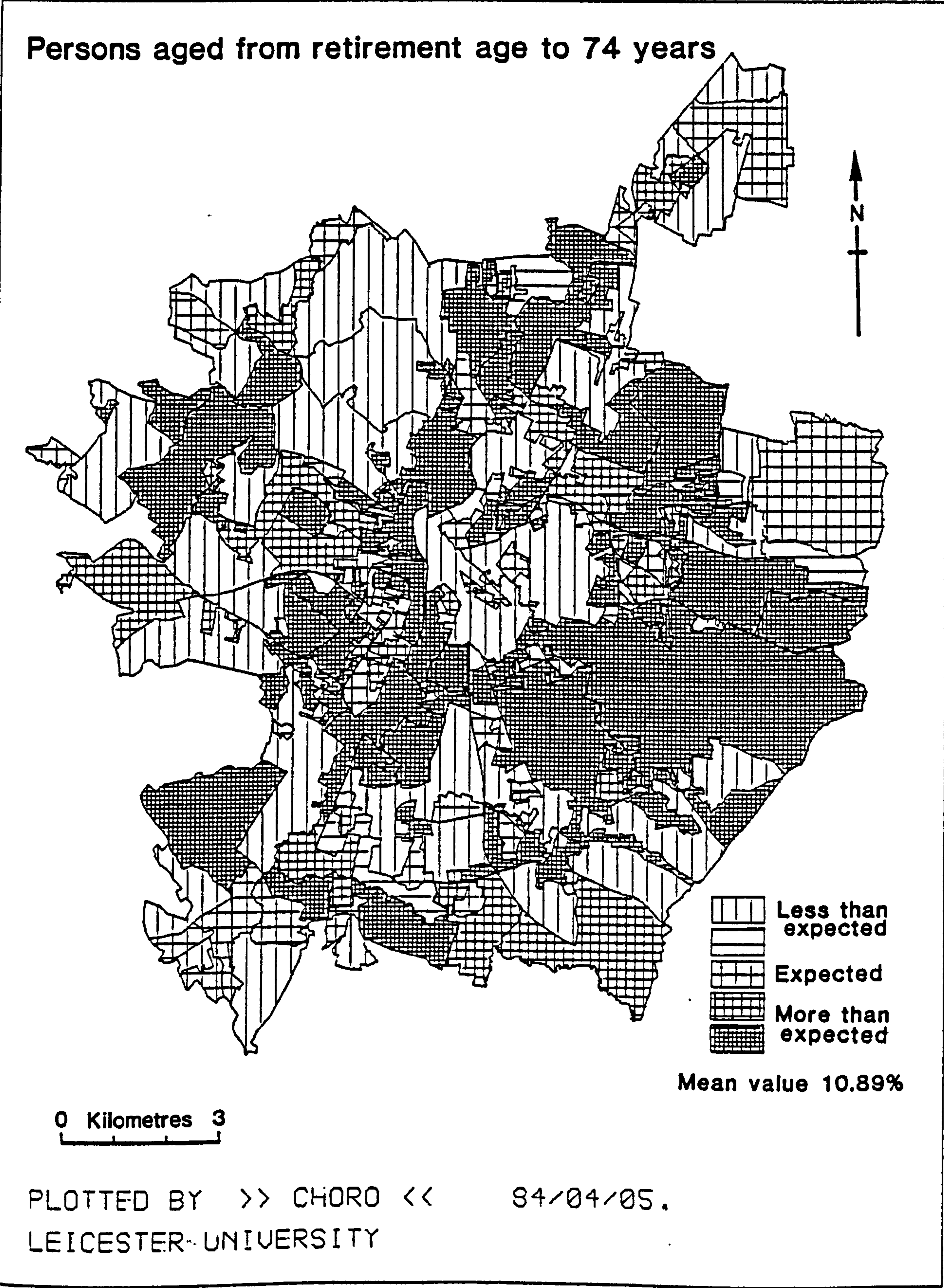
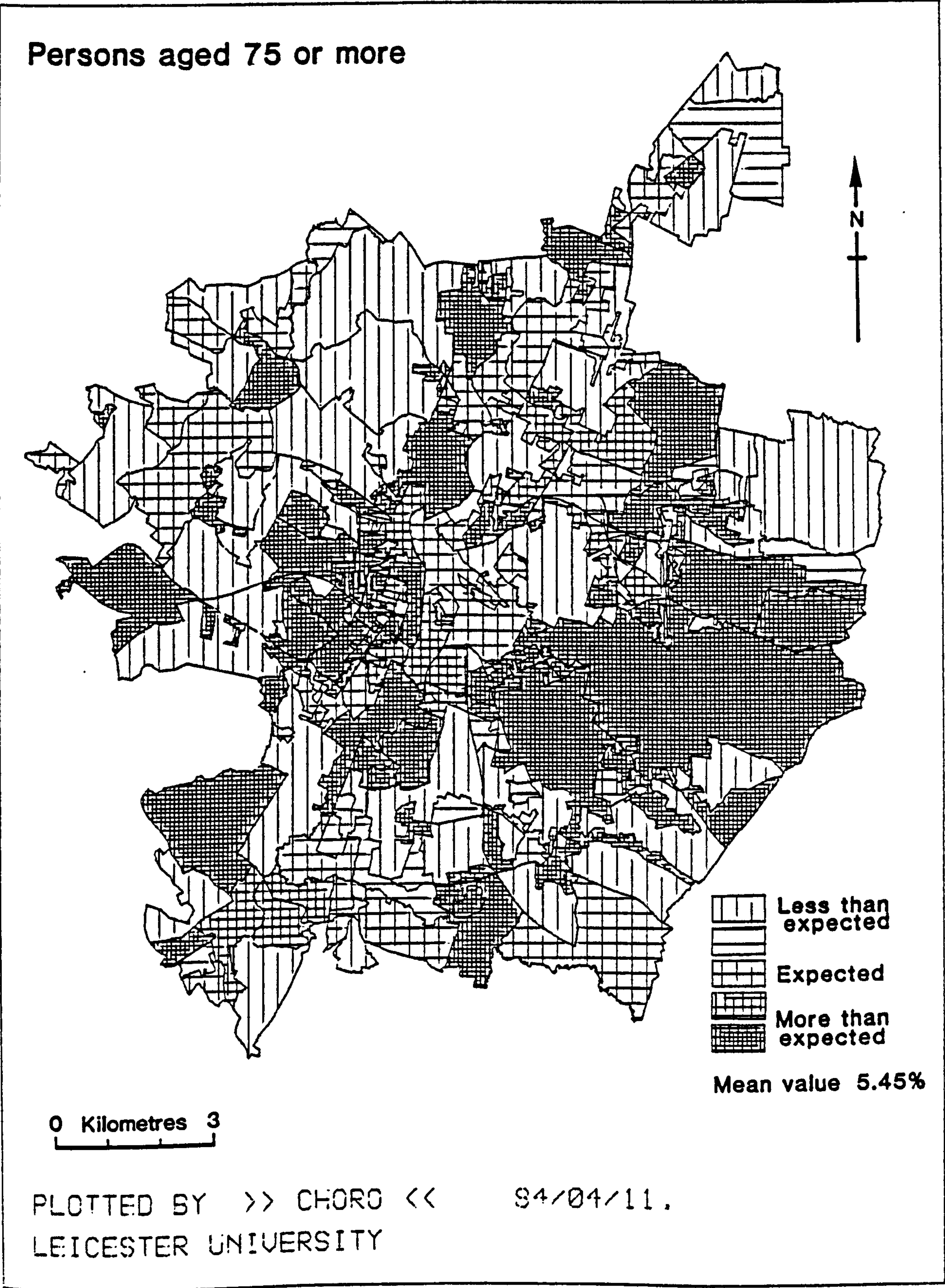




Figure 6.9





child (mixed households).

#### **Persons born in the New Commonwealth or Pakistan (Figure 6.10).**

At the time of the 1981 Census over 10% of the population of Greater Leicester were born in the New Commonwealth or Pakistan and over 20% of the city population. These figures do not of course include second and subsequent generations of Asians and therefore if anything understate the size and importance of the ethnic minority community in Leicester.

Spatially segregated, the Asians form a distinct cluster to the north east of the city centre in the Spinney Hill, Highfields, Belgrave and Rushey Mead areas; in one NSA (see Table 6.3) they comprise more than half of the population. A separate concentration is to be found to the west of the city centre in West End. Elsewhere in the study area disproportionately few people originate from the New Commonwealth or Pakistan.

The formation of distinct cultural subgroups represents a further stage in the development of the immigrant community. Within the longer established area of Highfields there are approximately equal proportions of Gujarati-speaking Hindus and Muslims and a diversity of other language and religion groups. Punjabi-speaking Sikhs are found in significant numbers at the edges of this area. Within Belgrave and Rushey Mead Gujarati-speaking Hindus predominate, in several places out-numbering the white population. (Leicester City Council, 1983). The evolution of the immigrant community has had and will have an important impact on the socio-economic geography of Leicester, influencing housing, retailing and industry.

#### **Single person households (Figure 6.11)**

Overall about 21.6% of households contain only one person, although within individual NSAs this proportion varies from 2.6%

Figure 6.10

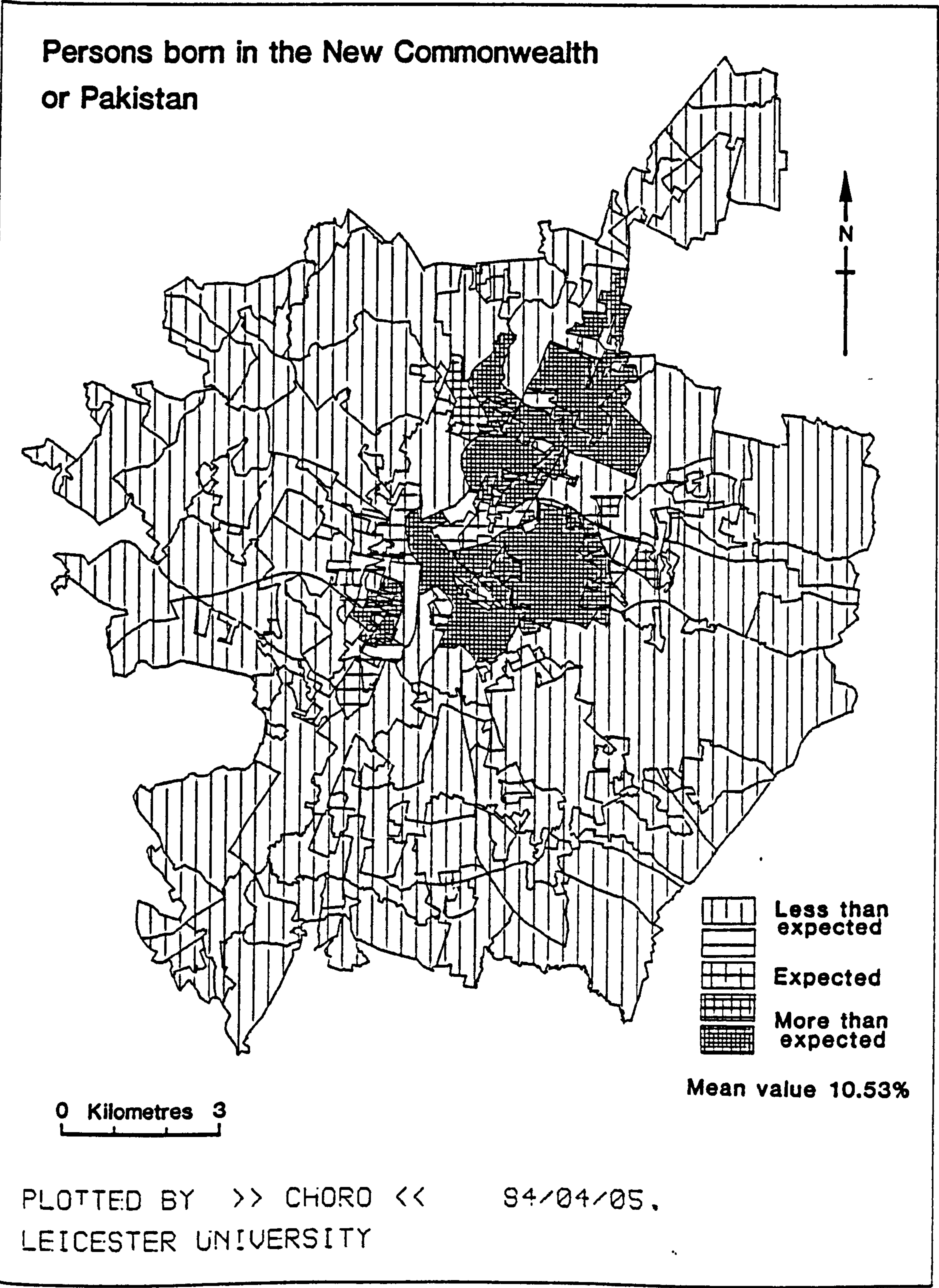
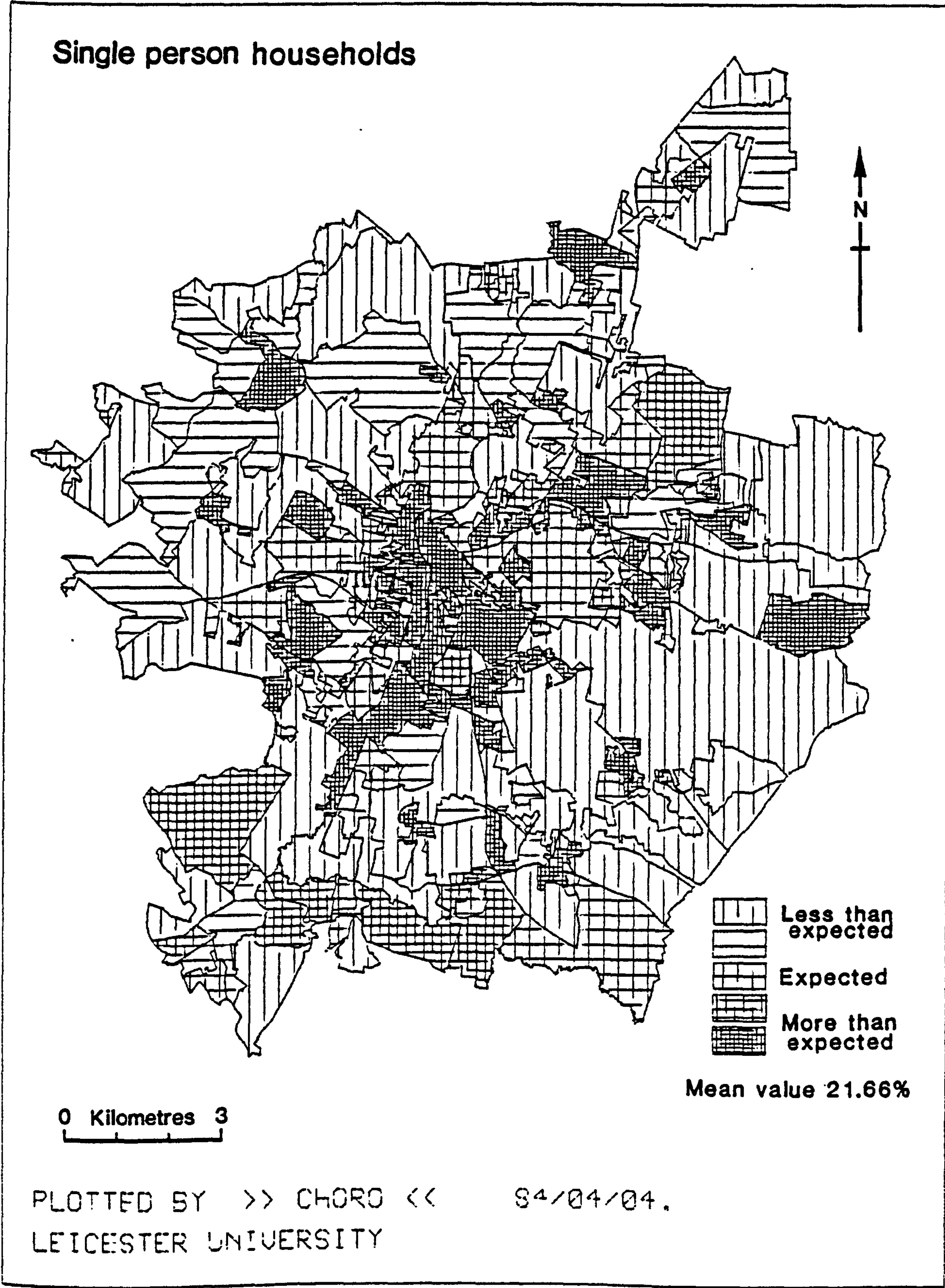




Figure 6.11





to 69.9%, a much broader range than found among the demographic variables (see Table 6.3). Spatially, single person households are concentrated in and near the city centre with additional enclaves at the older village centres, such as Wigston, Anstey, Thurnby and Braunstone.

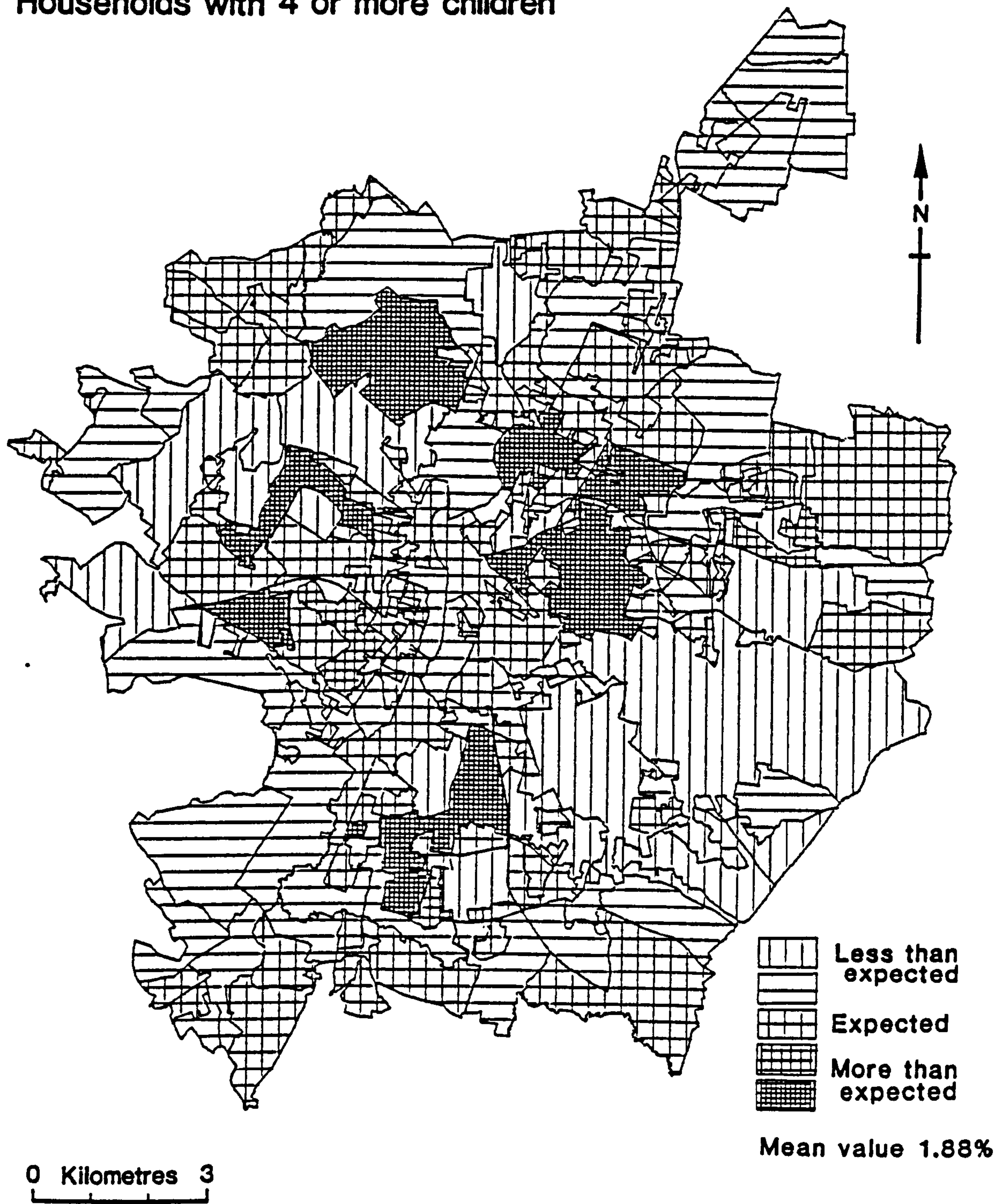
There are two main categories of single person household, differentiated by age. Firstly, there are households comprising young single people. These are predominant in the city centre where privately rented accommodation is readily available and the benefits of accessibility to work and entertainment are attractive forces. The second category comprises the elderly population, in particular widows and widowers. Within Greater Leicester these people are most frequently found in or near the city centre, in the older villages and in certain council estates. The distribution of single person households is a combination of the distributions of young adults (see Figure 6.4) and pensioners (see Figures 6.8 and 6.9); where the two reinforce each other extreme values are obtained, such as in the city centre, where they conflict an 'average' value results, for example around the inner suburbs.

#### **Households with four or more children (Figure 6.12)**

Areas having high proportions of large families include Highfields and Belgrave, due to the presence of the Asian community; the older council estates at Braunstone, New Parks, Saffron Lane and Eyres Monsell, each having fairly high numbers of larger properties; and the newer council estate at Beaumont Leys. Recent council policy has attempted to rectify the mismatch between housing and family size by rehousing families in larger properties. A higher percentage of four and five bedroomed houses have been built in the last twenty years than ever before. Beaumont Leys has a total of 128 post-1964 four and five bedroomed

Figure 6.12

Households with 4 or more children



PLOTTED BY >> CHORO << 84/04/04.  
LEICESTER UNIVERSITY



houses, while some of the older council estates, such as Saffron Lane, have recently had larger properties added to their existing housing stock.

Several NSAs have no households in this category at all. This may be due to a youthful age structure, a lack of suitable accommodation or to a general desire to maintain a standard of living.

#### **Lone parent households (Figure 6.13)**

Of all households in which children are present, over 6% contain only one adult. The distribution of these is in some respects similar to that of large families (see Figure 6.12), no doubt because the two groups are both favoured by council house allocation policy. The main differences lie in places such as Highfields and Belgrave which have few single parent families, probably for cultural reasons; and the city centre and Thurnby which have a relatively high proportion. Within the city centre the availability of small and cheap privately rented accommodation would be significant, as would access to employment, health care and other services.

#### **Mixed households (Figure 6.14)**

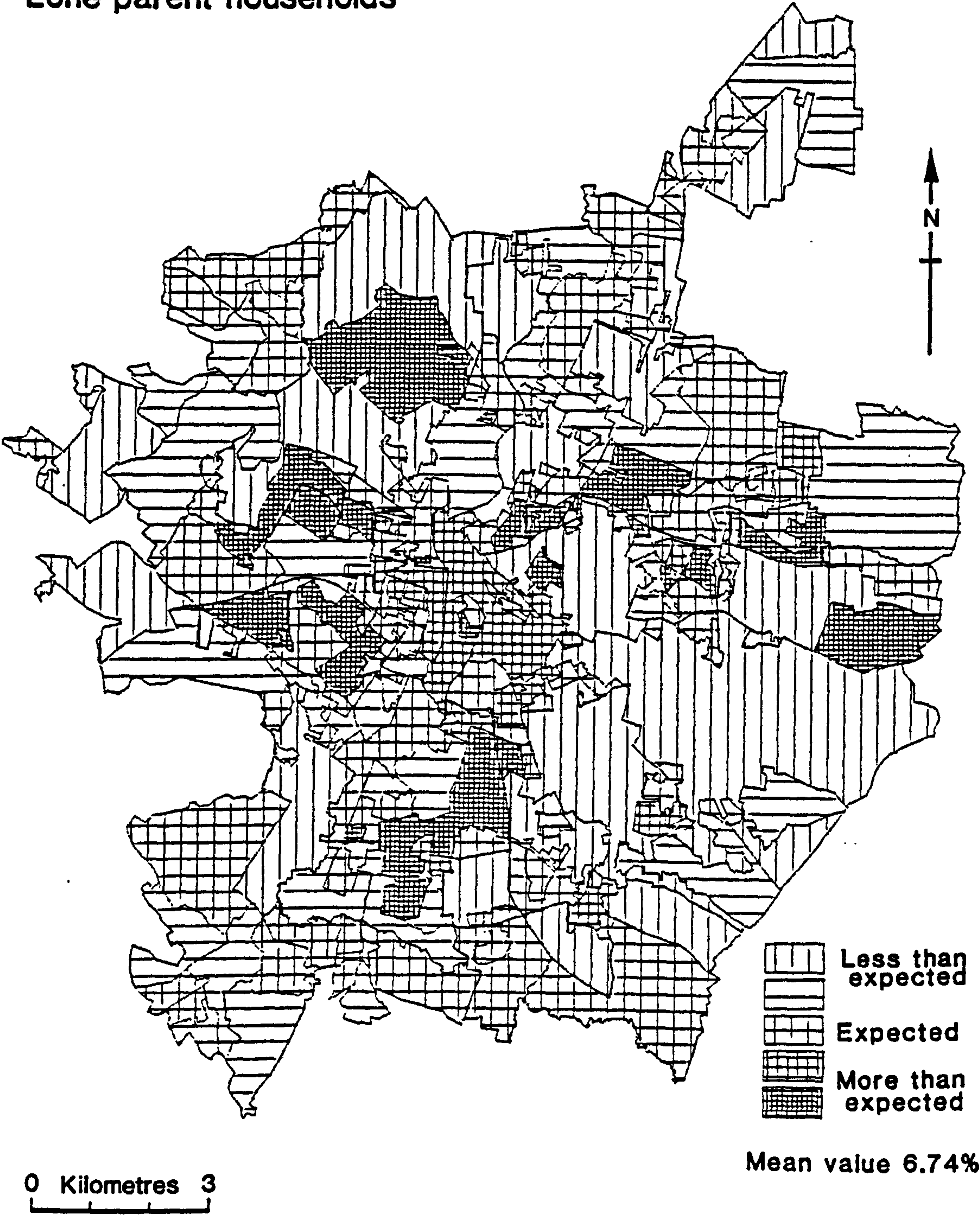
Mixed households refer to those containing three or more adults and one or more children aged less than 16. The household labelled thus might therefore consist of either the conventional British nuclear family in which the 'children' are both less than, and older than 16; or perhaps an 'extended' family in which additional adults, either related or not, are members of the same household.

Spatially, the distribution of areas with high proportions of mixed households is similar to the distributions of large and lone parent families (see Figures 6.12 and 6.13). The presence of



Figure 6.13

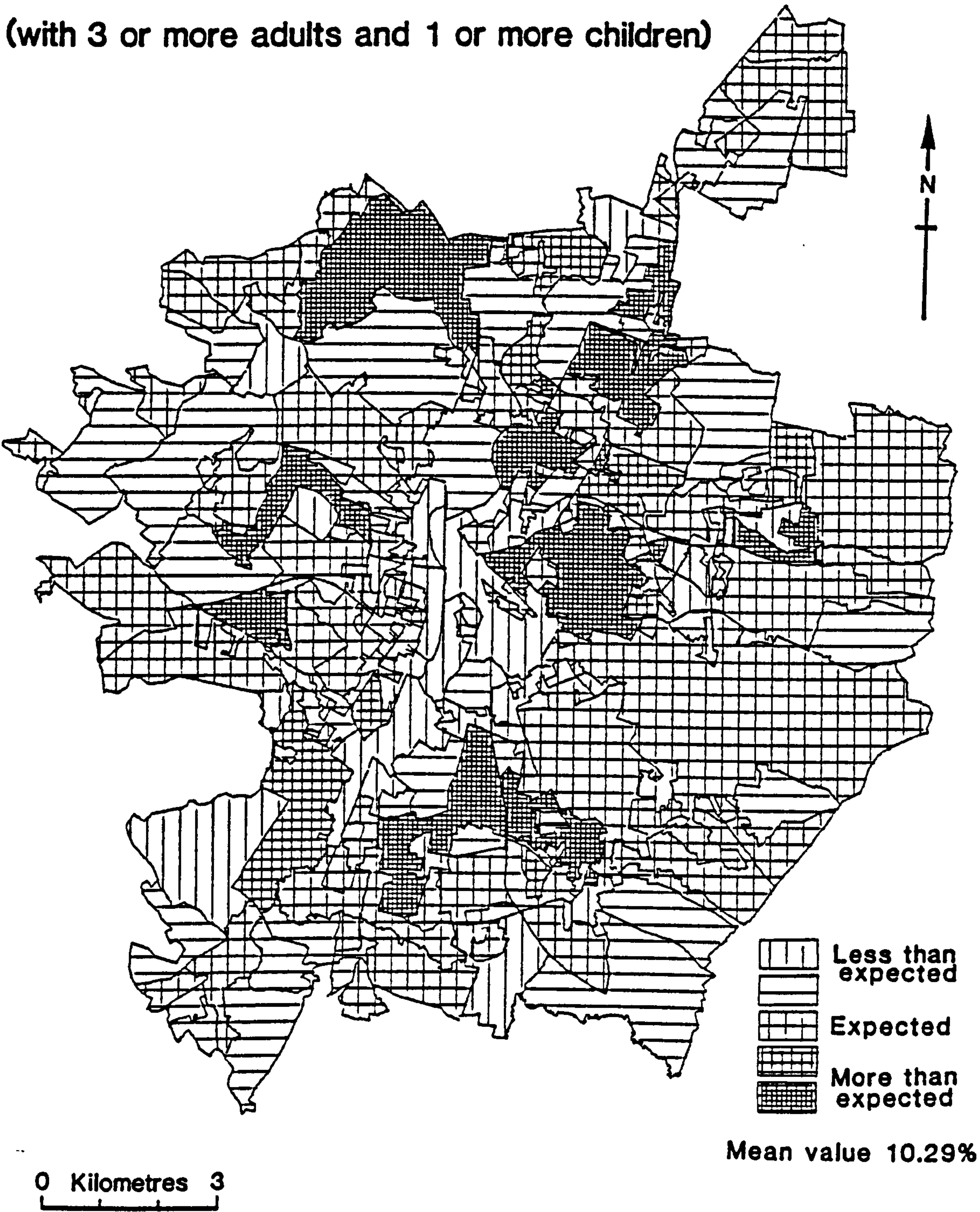
Lone parent households



PLOTTED BY >> CHORO << 94/04/04.  
LEICESTER UNIVERSITY

Figure 6.14

Mixed households  
(with 3 or more adults and 1 or more children)



PLOTTED BY >> CHORO << S4/04/04.  
LEICESTER UNIVERSITY



grown-up children in the family home probably accounts for the high proportions of mixed households in some of the older council estates, especially in Braunstone, New Parks, Saffron Lane and Eyres Monseil. Within the densely shaded ethnic community the 'extended' family is likely to be more common, a maximum value for mixed households of 33.5% of all households is achieved by one NSA within the Asian quarter.

In summary, the distribution of household types is influenced most strongly by three factors: life cycle stage, the distribution of the Asian and West Indian communities and that of council housing. Collectively these determine both the demand and supply of housing and have a major influence on the social geography of Leicester.

### 6.3.3 Housing and Tenure

Reference has already been made to housing tenure in describing the distribution of previous variables, this is because the availability of accommodation plays a prominent role in determining the social structure of the urban area. Housing is therefore important both to a description of the urban environment and to the identification of problem areas. Three tenure variables were included in the analysis: owner-occupied, council rented and privately rented property. In addition, four other variables were chosen to describe housing conditions, three of these being conventional indicators: households lacking full amenities, 'overcrowded' households, households not in self-contained accommodation, and finally, purpose-built flats.

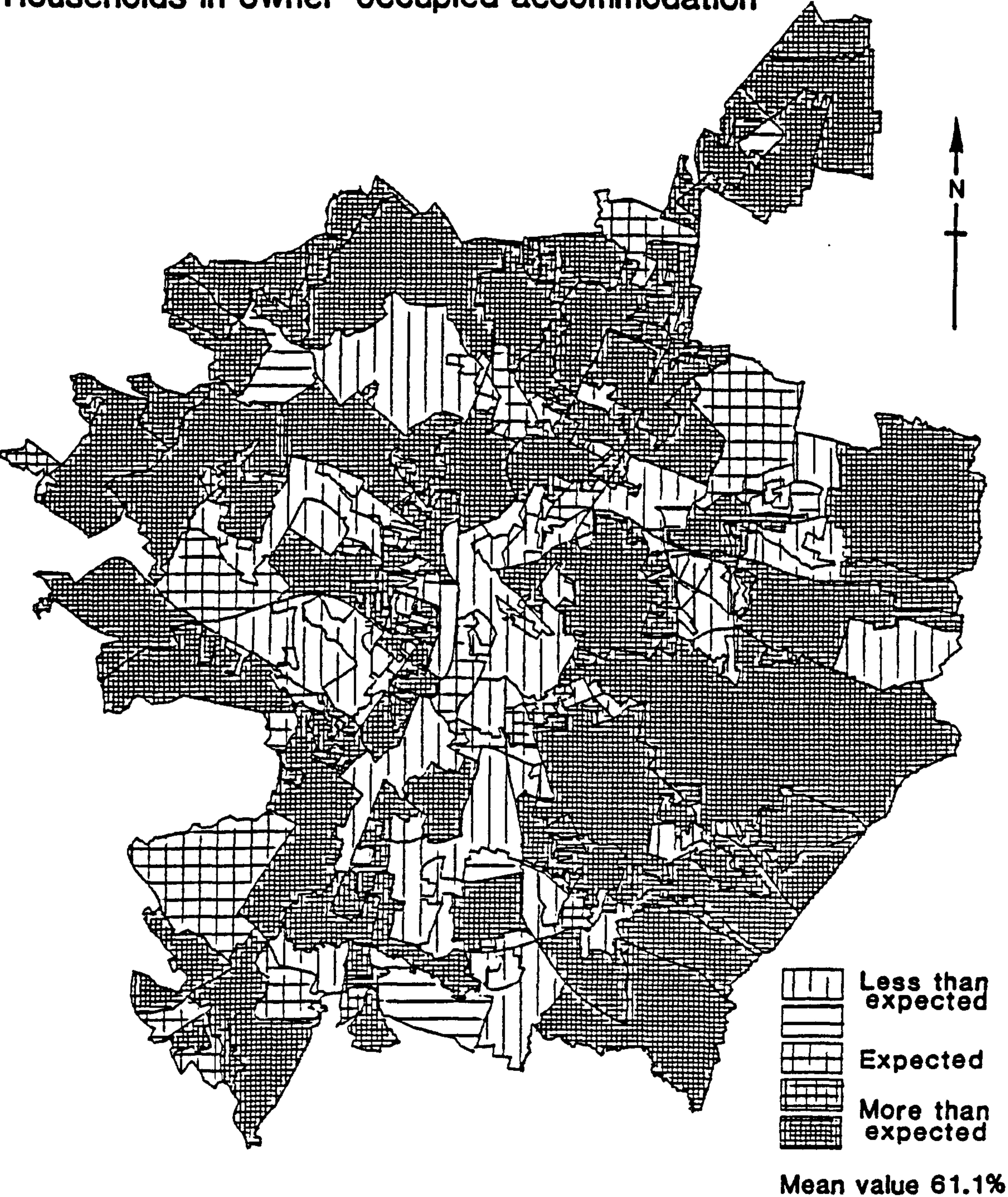
**Tenure** (Figures 6.15, 6.16 and 6.17)

Owner-occupied, council-rented and privately rented property between them comprise over 96% of all housing in Greater Leicester; the other tenure types (ie. housing association, tied accommodation and non-permanent dwellings) account for the



Figure 6.15

Households in owner-occupied accommodation



PLOTTED BY >> CHORO << 84/04/17.  
LEICESTER UNIVERSITY



Figure 6.16

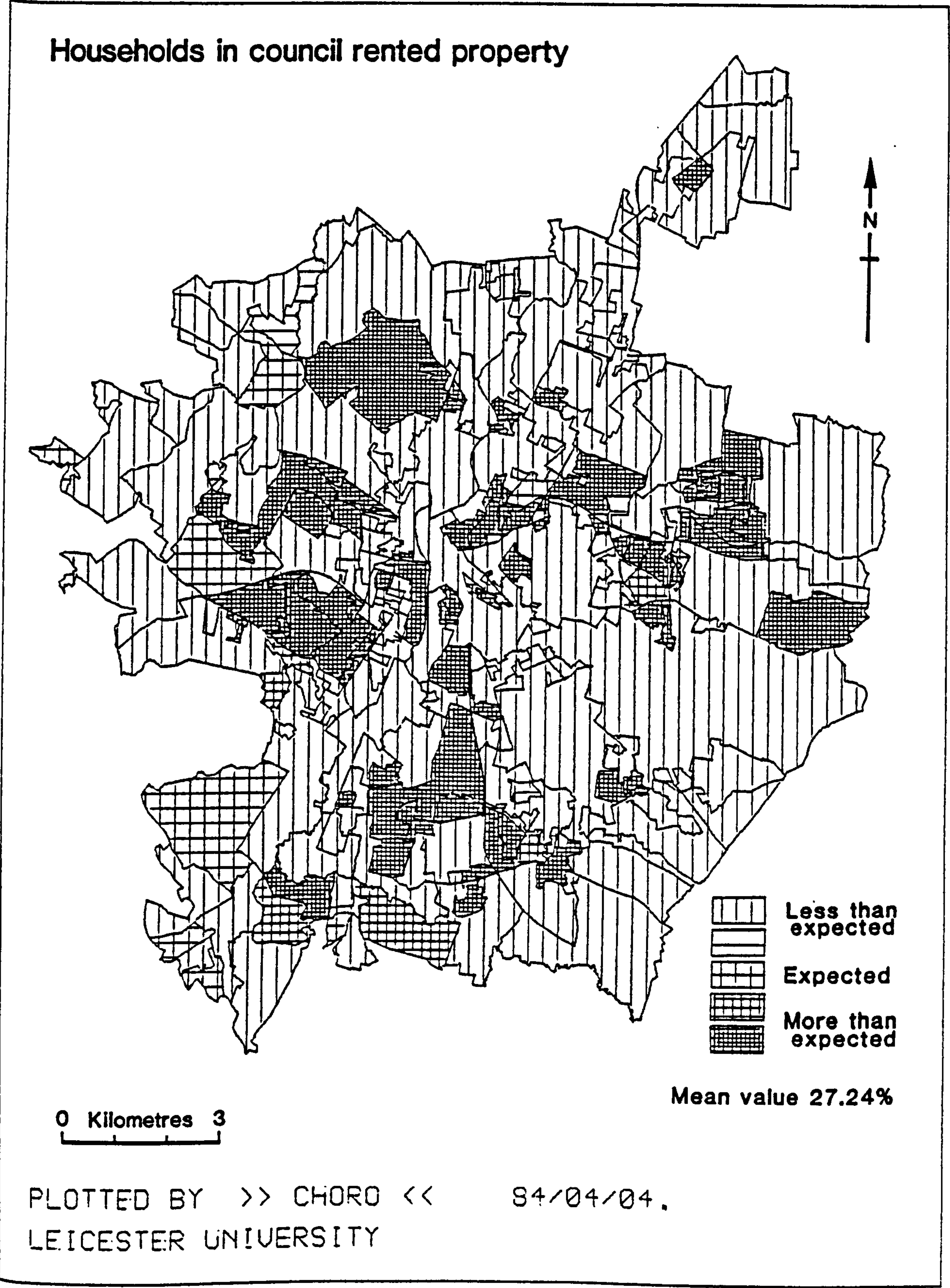
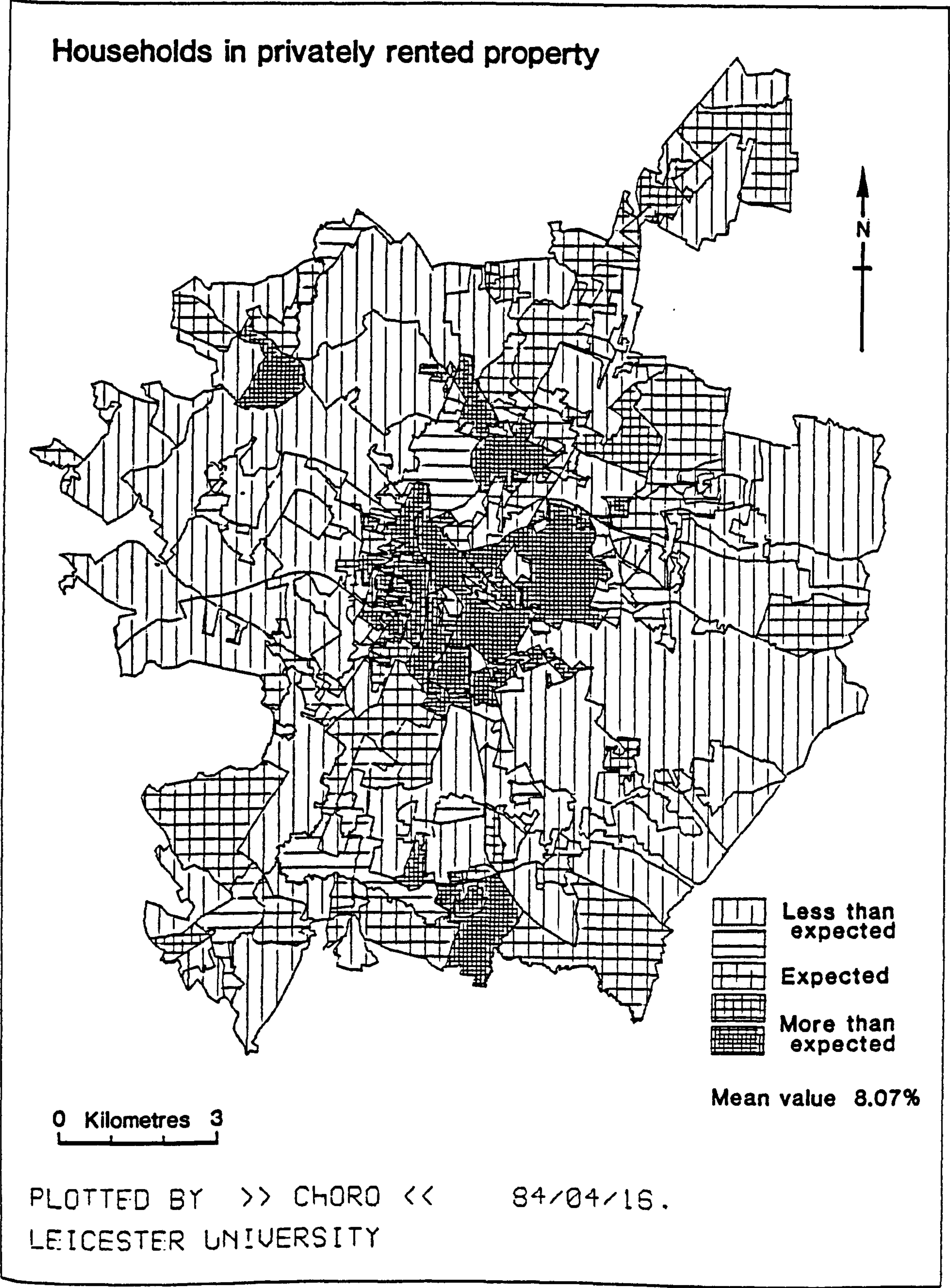


Figure 6.17.





remaining 4% but will not be considered here.

Figure 6.15 shows the predominance in Greater Leicester of owner-occupied housing, particularly in the outer suburbs. As mentioned previously, the majority of private housing development has always occurred in peripheral areas, this explains much of the dense shading outside the city boundary.

Within the city three related factors are significant in explaining tenure patterns. First the growth of Leicester has historically been greater to the east of the river, only from the late nineteenth century did residential development take place on the floodplain to the west. Thus the majority of the older working class property is now to be found to the north and east of the city centre, while the older high status housing forms a southeasterly wedge. Both areas are currently predominantly owner-occupied; the former by the Asian population, originally attracted to the Highfields area by the plentiful supply of cheap terraced housing; and the latter by the wealthier members of the community who are able to afford the large, low density homes.

Elsewhere in the city privately owned housing estates are interspersed with council housing. The earliest council estates to be occupied in Leicester were Northfields (1921) and Coleman Road (1922), the first building at Saffron Lane, Tailby and Braunstone followed soon after. Many of the post-war estates originate in plans devised in the 1920s and 1930s, and are built on land bought during this period (Evans, 1972). Like private development, the availability of land has always been a major constraint, with the largest estates only possible on the cheaper land at the edge of the city. For example, Braunstone has 5070 dwellings in total, New Parks has 3883 and Eyres Monsell, 3550. The newest major council estate, Beaumont Leys, had 2129 dwellings in 1982 and was still

expanding.

Recently, redevelopment of the inner area has provided the opportunity for building several smaller estates, these can be seen near the city centre on Figure 6.16. St. Andrews, St. Peters and St. Marks were all completed in the 1970s.

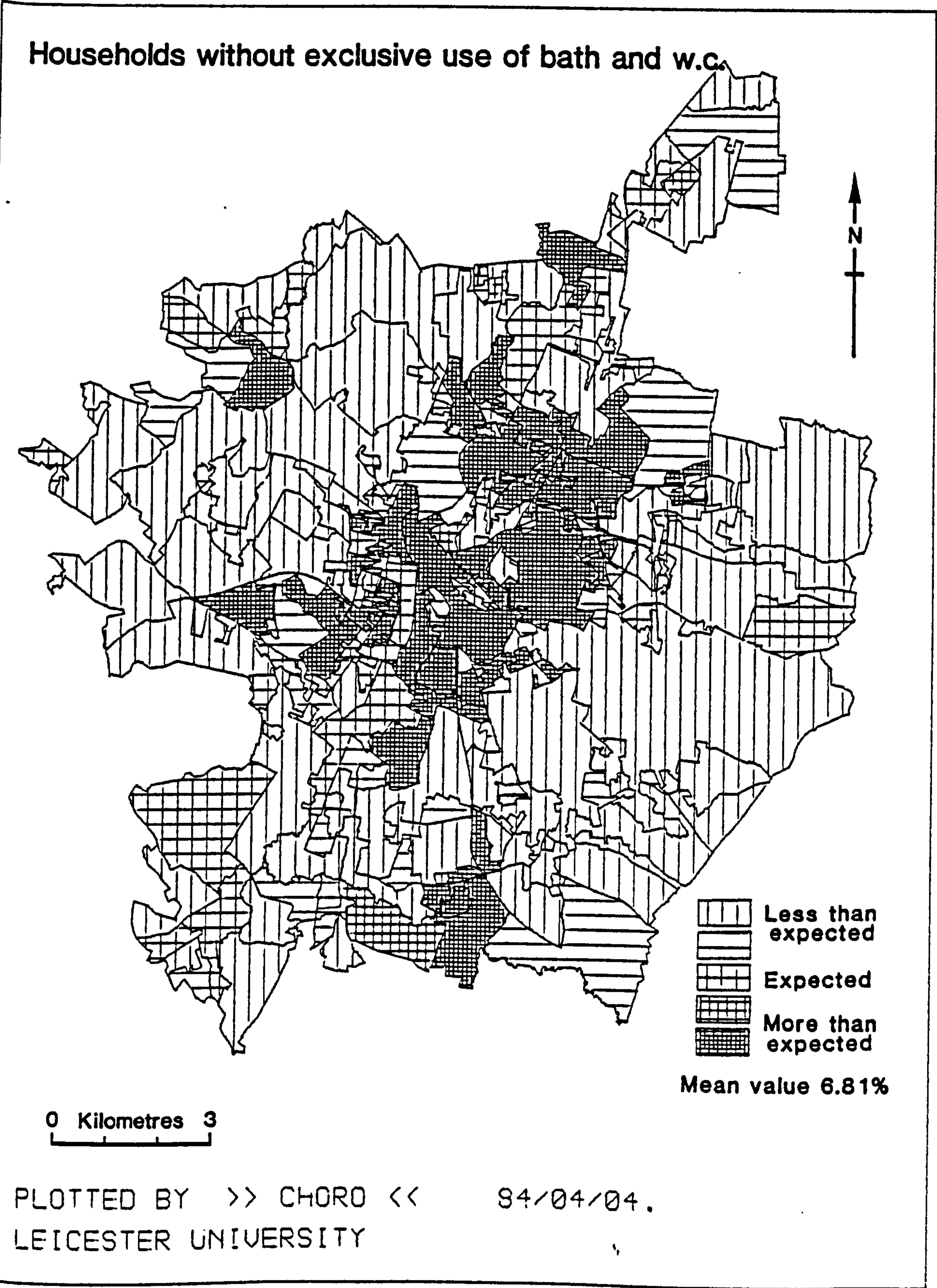
The third tenure type, privately rented accommodation, follows the classic British model. Areas with the highest proportions of this type of property include the city centre and the villages of Anstey and South Wigston. In each of these there is a predominance of nineteenth century property. Although the highest percentage of households in private property is as much as 79% (see Table 6.3), this value is exceptional; most of the areas to be highlighted by Figure 6.17 have about 20% privately renting households. This reflects the general trend in Britain away from private tenancy as both personal preferences and government legislation favour the other tenure types.

#### **Households lacking full amenities (Figure 6.18)**

'Amenities' in this context refers to the exclusive use of both a bath and inside W.C. Despite a substantial improvement over the last ten years in the provision of these facilities, Leicester still has a greater problem than many other major cities. Some 6.8% of all households in Greater Leicester are without sole use of amenities. Figure 6.18 suggests that this is a problem confined mainly to the city itself, and in particular to the areas of older private and council housing stock. Certain population groups are especially disadvantaged with respect to the availability of amenities. Private tenants and the ethnic minority groups not only tend to occupy the older and deteriorating nineteenth century housing, but also frequently share accommodation. Some of the early inter-war council estates such as Braunstone and Northfields also have high proportions of



Figure 6.18





households without amenities. Outside the city boundary the older villages of Anstey and Wigston are highlighted.

The City Council has been active in encouraging the improvement of older houses, having approved a Renewal Strategy and designated areas for priority treatment. In General Improvement Areas (GIAs) grants of up to 60% of improvement costs are normally available, while in Housing Action Areas (HAAs) up to 75% or 90% may be awarded. The success of these policies is illustrated by the drop in the proportion of households in the city which lack an inside W.C. from 27.9% in 1971 to 7.1% in 1981, and corresponding figures for those lacking a bath of 11.4% and 1.9% (Leicester City Council, 1983).

#### Overcrowding (Figure 6.19)

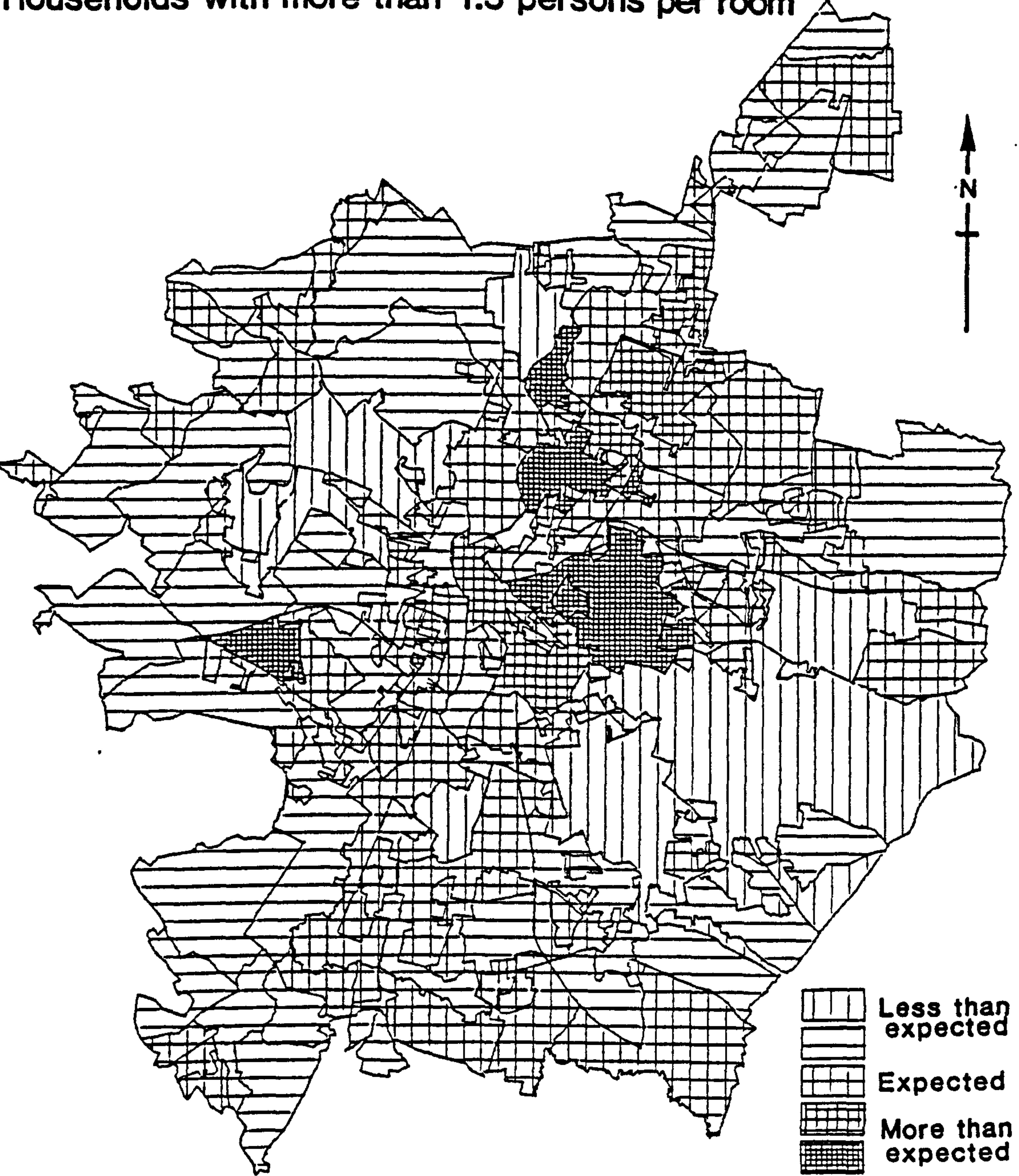
Overcrowding is measured here by the number of persons per room in a household; small kitchens, bathrooms and W.Cs normally being excluded. A critical value of 1.5 persons per room is conventionally taken as the lower limit for 'overcrowding', although some more recent studies have used a figure of 1.0, reflecting the nationwide improvement in housing standards.

Of all households in the study area, less than 1% live at densities of more than 1.5 persons per room. In many areas, particularly those of high socio-economic status, this figure is as low as 0%. The maximum value is 5.8%.

There are three main areas with a high proportion of overcrowded households, in each the problem is one of mismatch between family and housing size. Within Highfields and Belgrave the presence of the Asian community, having different expectations and cultural norms from the white population, accounts for the high densities. At Braunstone a different community is affected with the problem occurring in an area of predominantly council-

Figure 6.19

Households with more than 1.5 persons per room



Mean value 0.87%

PLOTTED BY >> CHORO << 84/04/04.  
LEICESTER UNIVERSITY



owned accommodation. The attempts by the City Council to relieve pressure in this area have already been mentioned (see Section 6.3.2).

#### **Households not in self-contained accommodation (Figure 6.20)**

As with the previous variable, very few people are 'deprived' in the sense of not living in self-contained accommodation, overall less than 1% of households fall into this category.

Figure 6.20 shows a very marked concentration towards the city centre, with the highest value of 39.5% occurring in the West End area. As noted before, the city centre is an area surrounded by both high density terraced housing and large Victorian properties. Much of this housing is suitable for subdivision and is made available for multiple tenancy. It is the typical 'bedsitter-land' common to most large British cities.

#### **Purpose-built flats (Figure 6.21)**

A distinction is drawn between the type of flat discussed above which has been created from the conversion of an existing older property, and the type that has been purpose-built as part of a local authority or private development. The latter is likely to be found anywhere in the urban area, depending on the availability of a suitable new or cleared site.

Almost 8% of household spaces in Greater Leicester are contained in purpose-built flats. These are predominantly council-owned, although some privately-owned flats are to be found in the city centre and to the immediate south east in Stoneysgate. Unlike many other major cities, a policy of building high-rise tower blocks was never adopted in Leicester (Evans, 1972) and most flats now are found in low-rise, single entry units.

The newer council estates tend to have proportionately more flats than those that are longer established. Beaumont Leys has



Figure 6.20

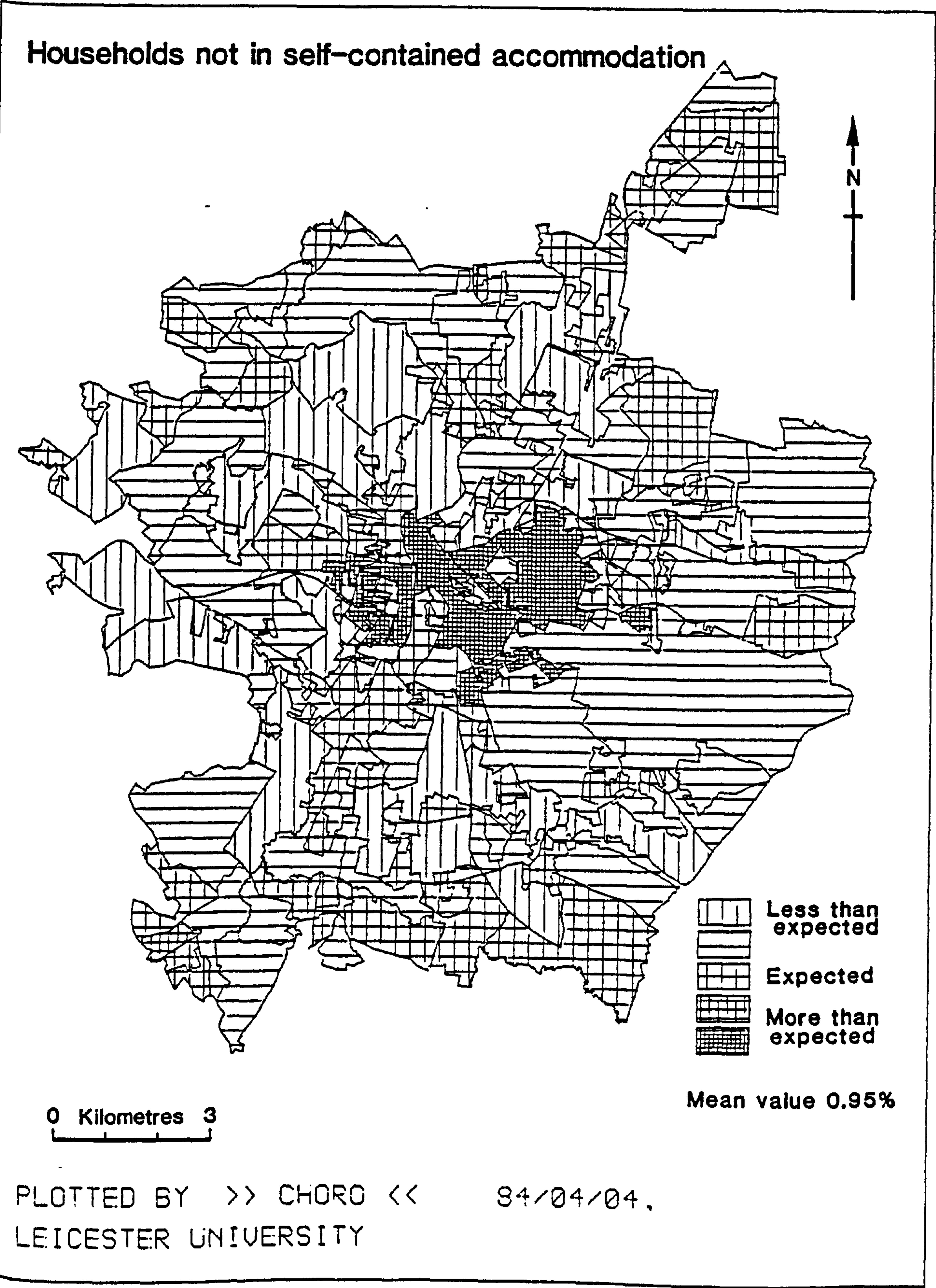
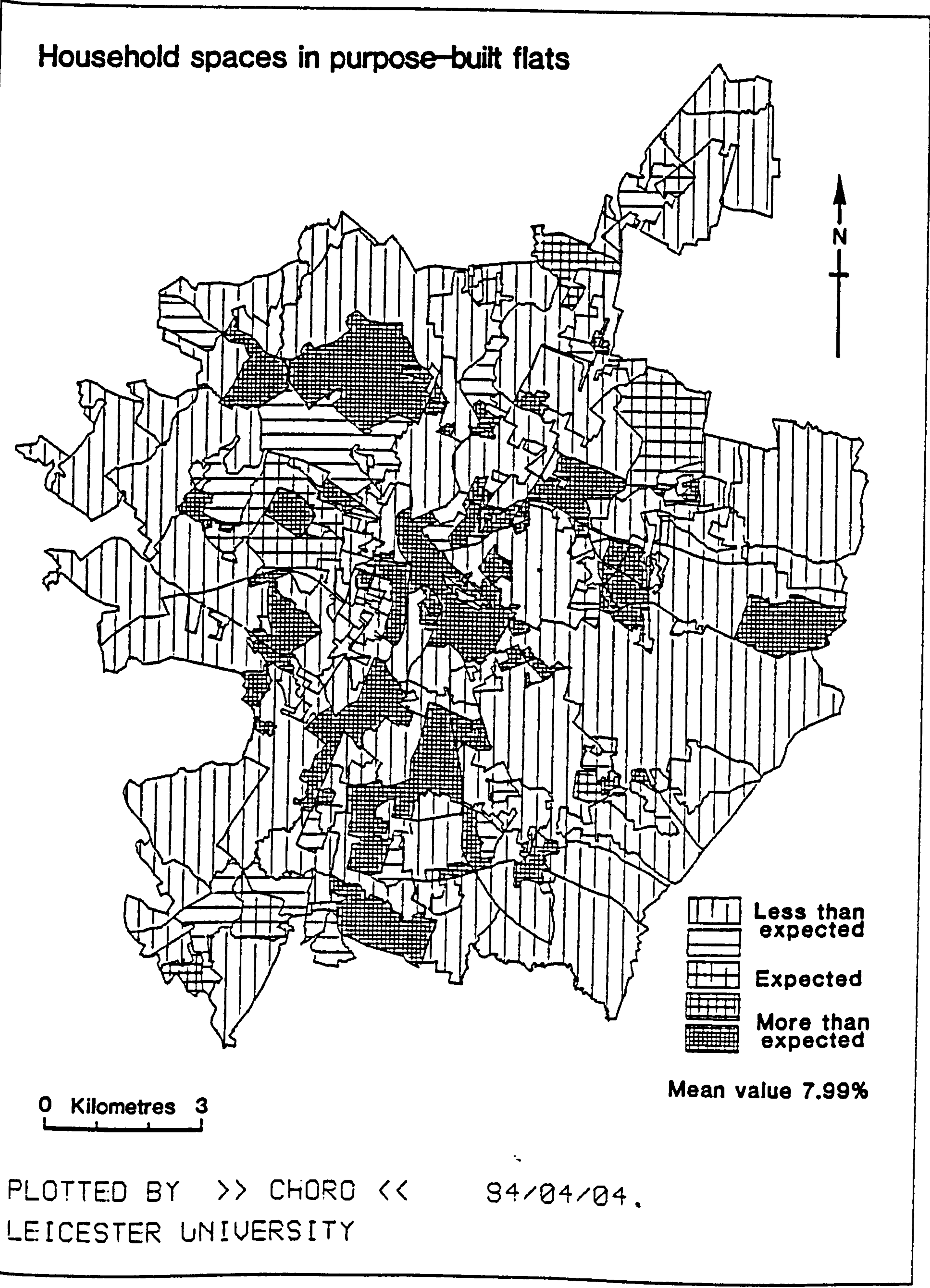


Figure 6.21





818 one- and two-bedroom flats, Rowlatts Hill has 539, St. Peters 436, St. Matthews 325 and Charnwood estate 316. Two factors account for this. First, there is an increasing demand for single person accommodation among the groups traditionally served by council housing, and second there is financial pressure to make the best use of redeveloped inner city sites. Among the older estates, Braunstone, New Parks and Eyres Monseil have significant numbers of purpose-built flats, with 766, 514 and 416 respectively. In Braunstone the majority are one-bedroom 'flatlets' occupied by senior citizens. Single people of any age have a greater chance of being allocated a council flat than a council house.

#### **6.3.4 Socio-economic status**

Following social survey convention five socio-economic groups were identified for this analysis, each containing one or more of the seventeen occupationally-derived classes defined by the census. One additional variable was included - adults with degrees, professional or vocational qualifications. It should be noted that all of these variables are based on the 10% sample data.

##### **Professional and managerial headed households (Figure 6.22)**

This category shows the widest variation across the study area, with proportions ranging from 0% in some small areas within Belgrave to 75% around Anstey. The average value for Greater Leicester as a whole is just over 15%.

Broadly speaking, the highest concentrations of this socio-economic group are found in the affluent outer suburbs, particularly in the south-east sector of Oadby, Stoughton, Stoneygate and Evington; other high status areas include Groby, Ratby, Glenfield and Kirby Muxloe to the west; Anstey, Thurmaston,



**Table 6.4 OPCS and Perinatal Mortality Survey (PMS) definitions**

**SOCIAL CLASSES**

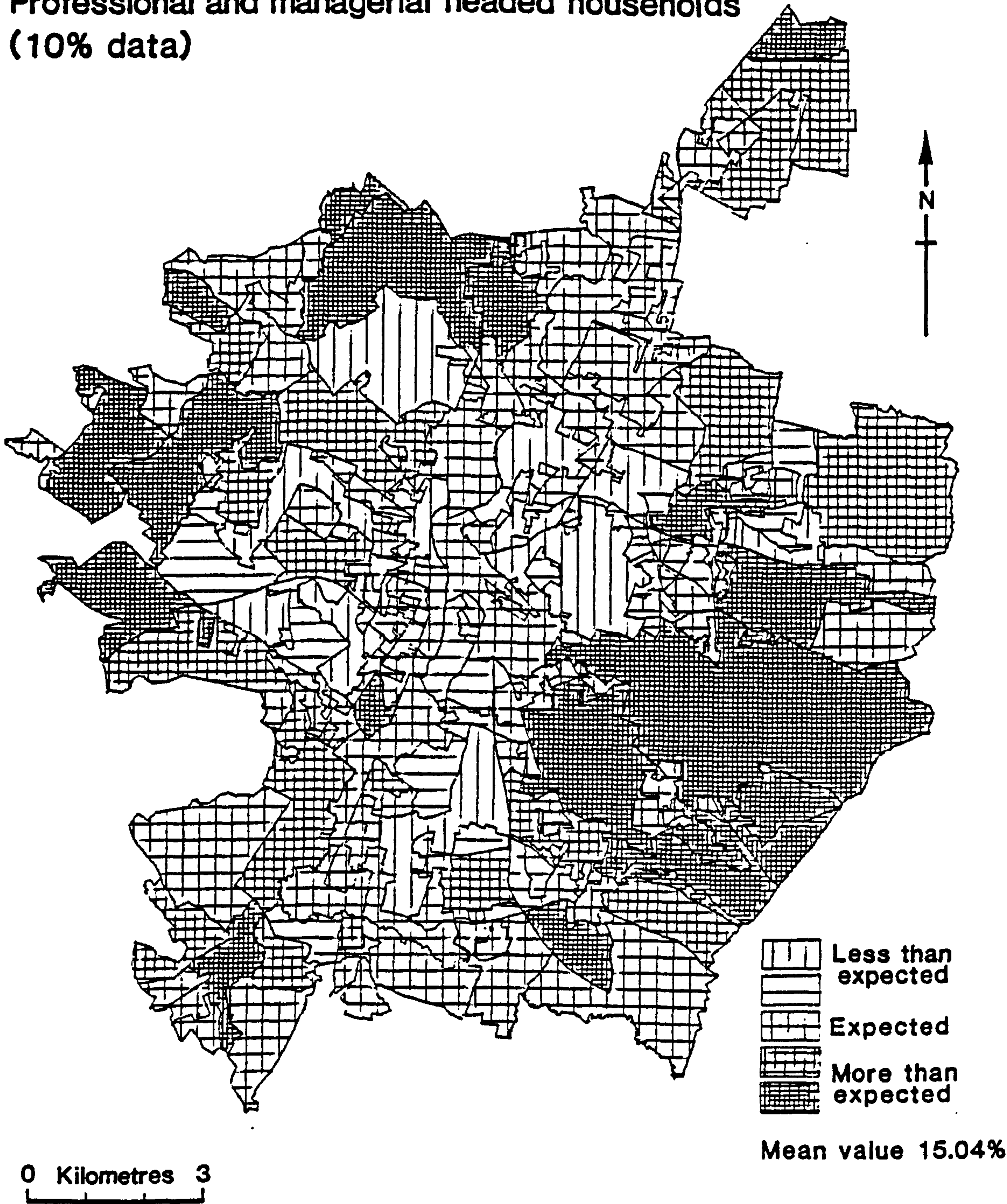
I	Professional etc. occupations
II	Intermediate occupations (generally including people of manager status)
III	Skilled occupations (manual and non-manual)
IV	Semi-skilled occupations
V	PMS - students
VII	PMS - housewives
0	PMS - armed forces

**SOCIO-ECONOMIC GROUPS**

1	Employers and managers in central and local government, industry, commerce etc. - large establishments
2	Employers and managers in industry, commerce etc. - small establishments
3	Professional workers - self-employed
4	Professional workers - employers
5	Intermediate non-manual workers
6	Junior non-manual workers
7	Personal service workers
8	Foremen and supervisors - manual
9	Skilled manual workers
10	Semi-skilled manual workers
11	Unskilled manual workers
12	Own account workers (other than professional)
13	Farmers - employers and managers
14	Farmers - own account
15	Agricultural workers
16	Members of armed forces
17	OPCS - inadequately described and not stated occupations
	PMS - housewives and unemployed fathers
77	PMS - students
88	PMS - not applicable

Figure 6.22

Professional and managerial headed households  
(10% data)



PLOTTED BY >> CHORO << 84/04/11.  
LEICESTER UNIVERSITY



Birstall and East Goscote to the north; and Wigston and Narborough to the south. Closer to the city centre, concentrations are much lower and there is no obvious sign of any high status enclave such as that found in many larger cities.

#### **Non-manual headed households (Figure 6.23)**

The distribution of non-manual headed households in the main reflects the pattern of the previous variable. The long-standing high status south east sector dominates the map, and many of the outer suburbs previously mentioned have above average proportions of non-manual households. In addition, the central area is densely shaded, accounted for by the presence of young people residing close to their workplace in the city centre; also heavily shaded is the popular late nineteenth century district of Western Park. The area with the highest proportion (61.5%) of non-manual headed households is to be found due south of the city centre, between Glen Parva and Wigston.

Few areas have significantly less non-manual headed households than expected, most notably the ethnic quarter and some of the older council estates.

#### **Skilled manual headed households (Figure 6.24)**

Almost 30% of households are headed by people in skilled manual occupations, making this the largest of the five groups. Figure 6.24 shows a distribution that is in complete contrast to the previous two maps. The south east sector and the north west fringe have proportionally fewer skilled manual headed households than the study area as a whole, while some of the inner suburbs have proportionally more. There is evidence of concentration in the older suburbs, particularly those established in the inter- and immediate post-war periods. This applies to both council and owner-occupied housing. For example the large council estate of New Parks was designed primarily for this sector of the population



Figure 6.23

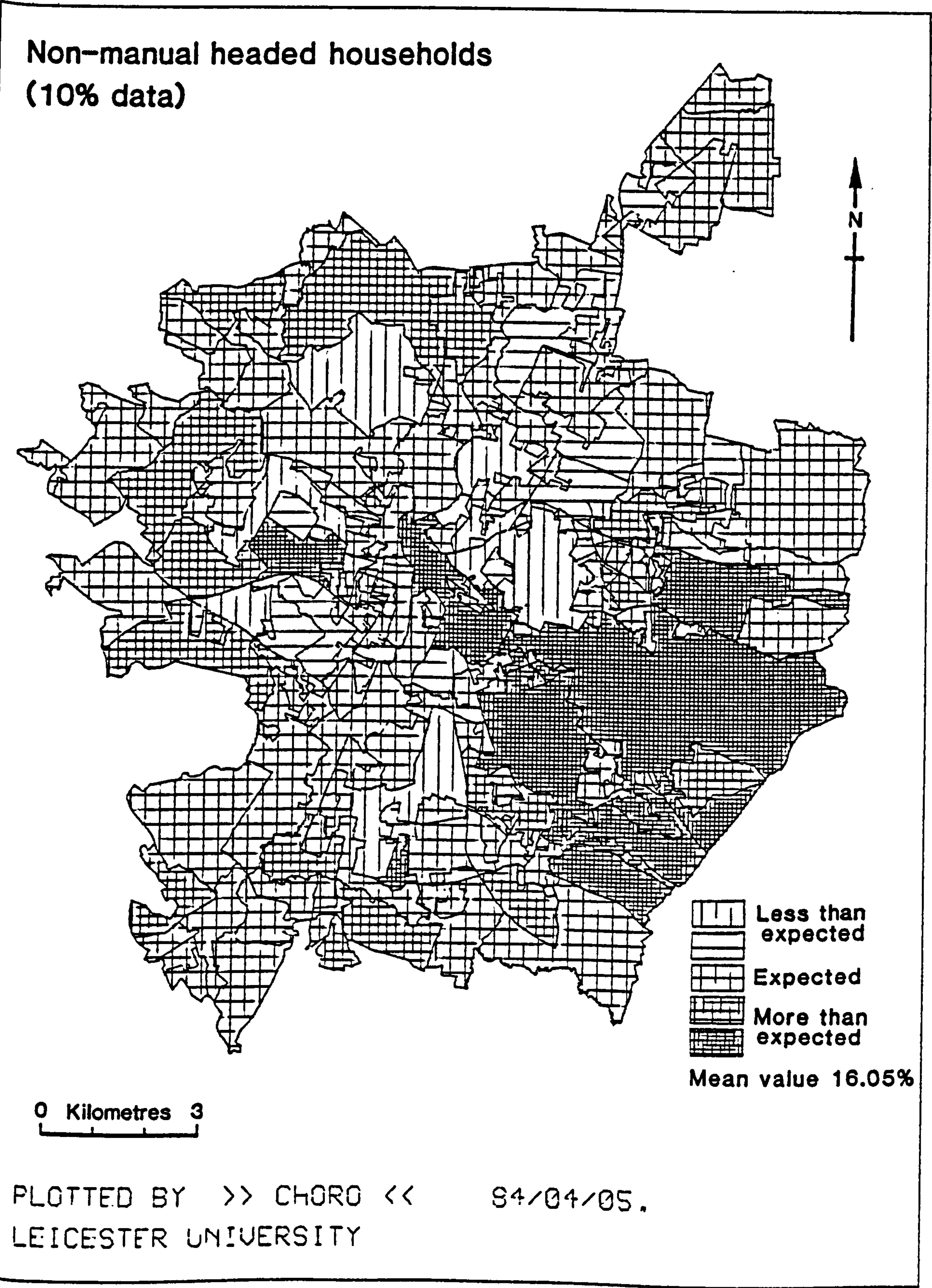
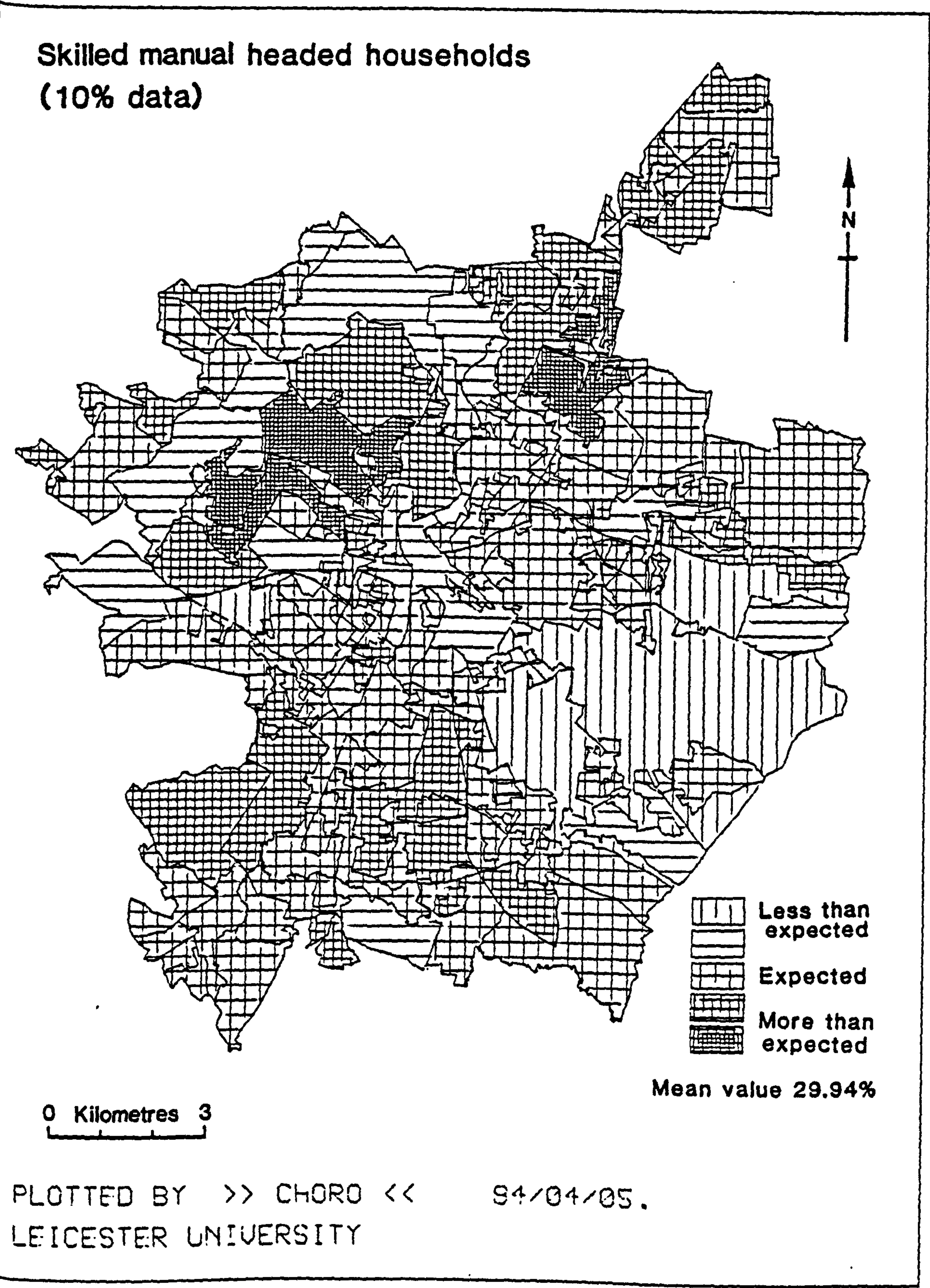




Figure 6.24



whilst more recently, private development at Rushey Mead has catered for owner-occupiers in this group. Glenfields and Thurmaston also have above average skilled manual households.

#### **Semi-skilled manual headed households (Figure 6.25)**

Five areas stand out as having high proportions of semi-skilled manual headed households. Two of these are council estates, New Parks and Beaumont Leys, the other three are older inner city areas, Highfields, Belgrave and the Meacham Road/Stadium area. As expected all of these are areas of low cost accommodation, in addition they are also home to the majority of the city's Asian and West Indian communities.

#### **Unskilled manual headed households (Figure 6.26)**

With less than 4% of households in Greater Leicester being headed by unskilled people, this is the smallest socio-economic group. The older council estates of New Parks, Braunstone and Eyres Monsell have high proportions of such households, as does Highfields - centre not only of the ethnic minority community but also home to many of the less able white population as well. Interestingly, with the exception of some of the particularly affluent suburbs, most of the study area has close to the average proportion of unskilled households.

#### **Adults with post school qualifications (Figure 6.27)**

Overall, some 5.5% of the adult population have degrees, professional or vocational qualifications. The greatest proportions of these people are found in the higher status areas (note the correlation between Figure 6.27 and Figures 6.22 and 6.23), particularly in the city centre and the south east sector. Two reasons may be given for this distribution. First, the better qualified people command higher salaries and eventually the more desirable property; secondly, there are an increasing number of



Figure 6.25

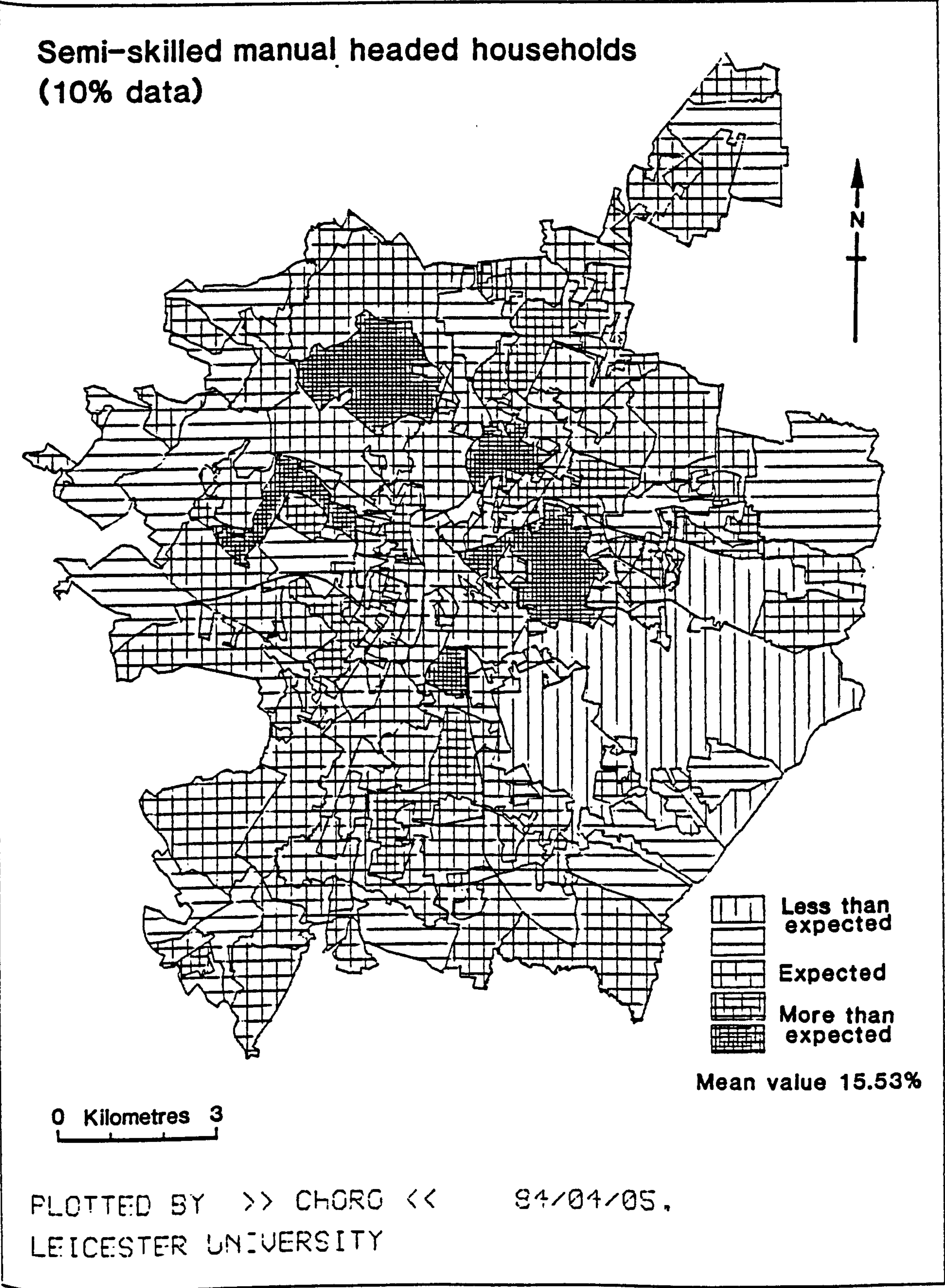


Figure 6.26

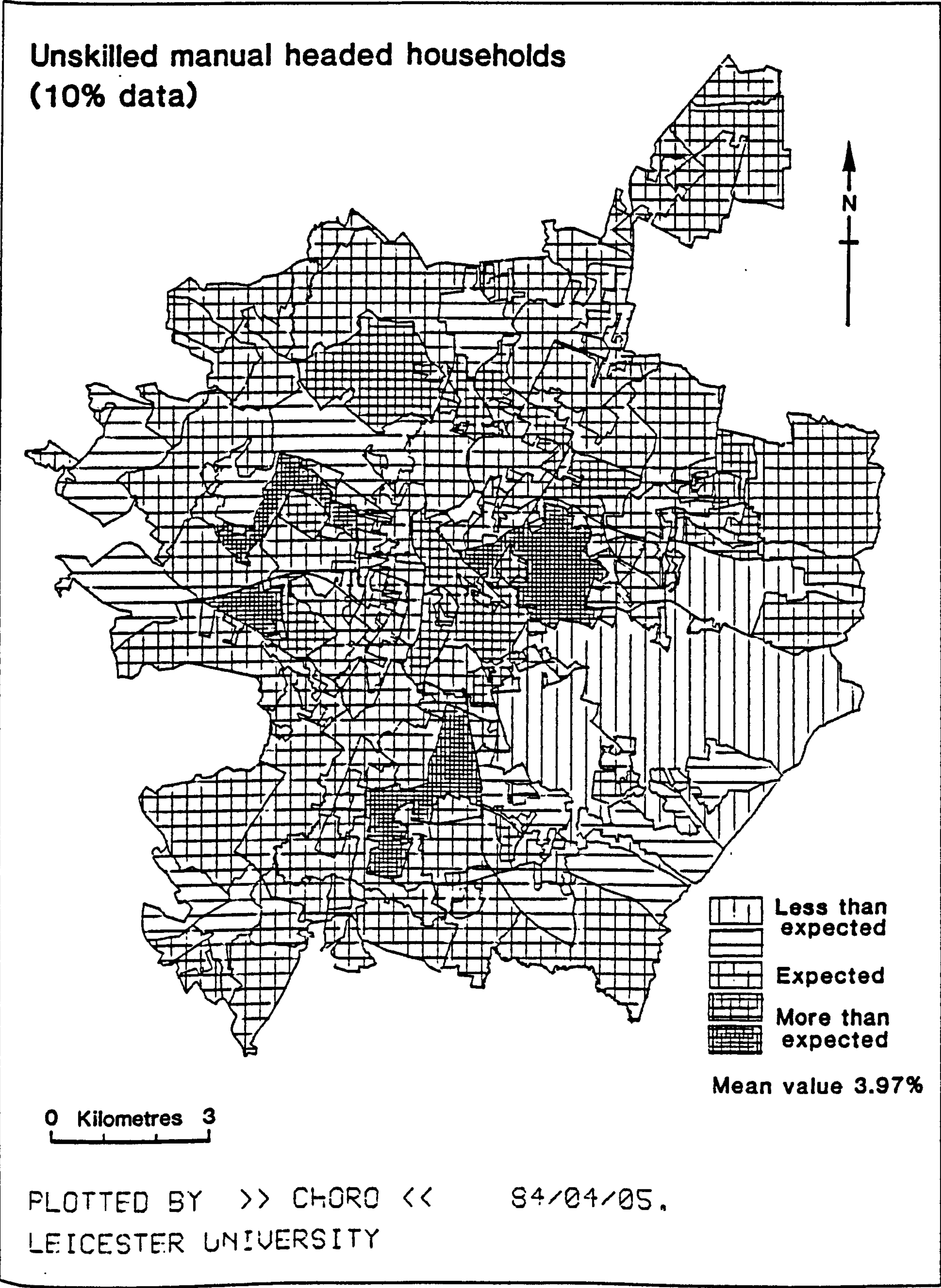
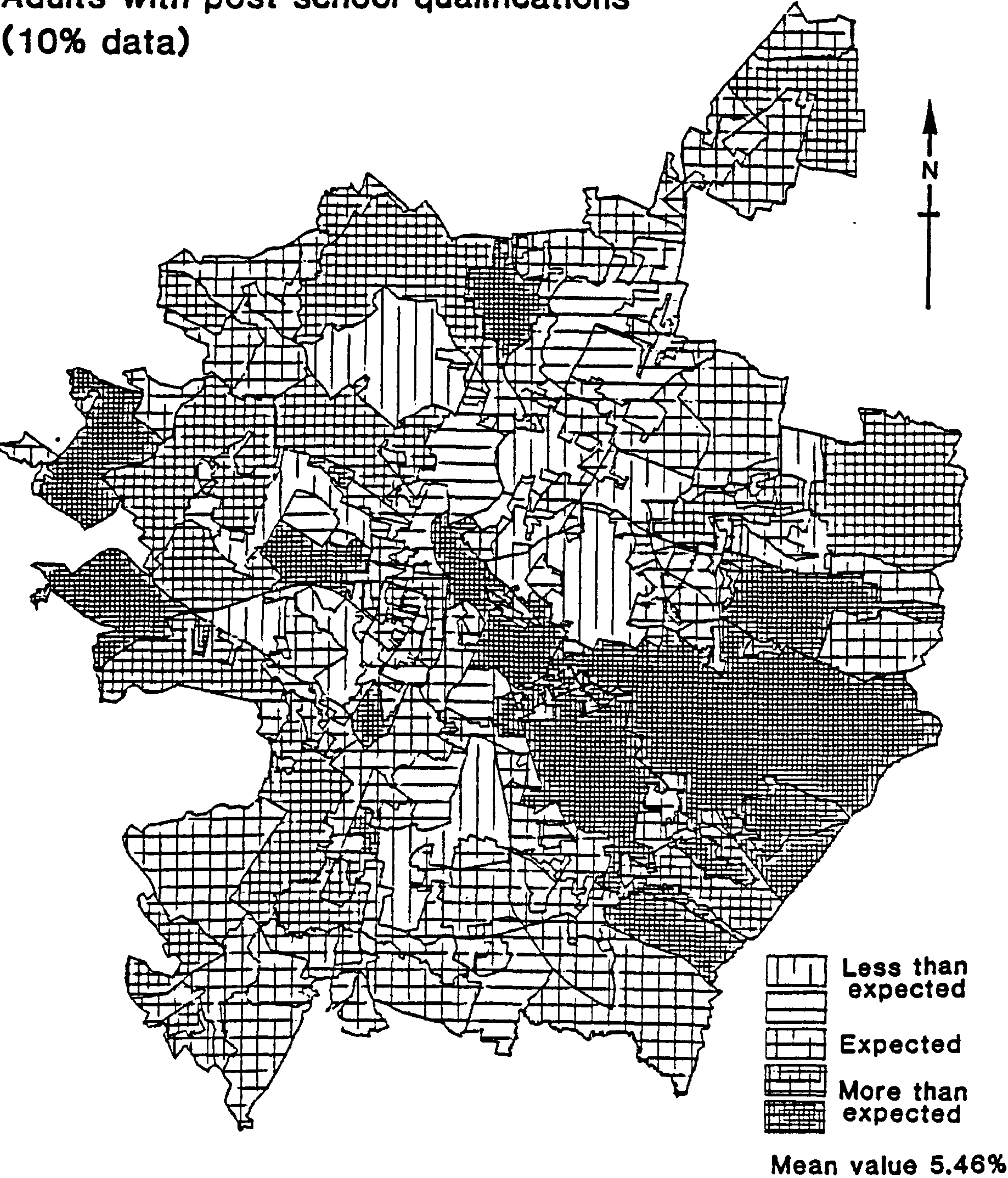




Figure 6.27

Adults with post 'school qualifications  
(10% data)



PLOTTED BY >> CHORO << 84/04/05.  
LEICESTER UNIVERSITY



younger people with post school qualifications who do not have the resources for this type of property and in the meantime search for cheaper temporary accommodation, often in or near the city centre.

At the moment there are markedly few adults with post school qualifications in the Asian areas of Highfields, Belgrave and Rushey Mead. This is likely to change as the ethnic community becomes more established (see Figure 6.31).

#### **6.3.5 Economic Activity**

Whereas socio-economic status describes the population with respect to occupation, levels of economic activity are perhaps more useful in identifying problem areas, particularly zones of high unemployment. According to OPCS, an economically active person is one who is either working, seeking work, prevented from seeking work by temporary sickness, or waiting to take up a job. Housewives, retired people, students and those who are permanently sick or disabled are excluded.

Seven variables in the present study concern themselves with the level of economic activity in the population: the number of economically active people aged over 16 (this serves as a base for comparison with the other groups), households with two or more such people, unemployed people, students, married women in employment, the young unemployed and self-employed people.

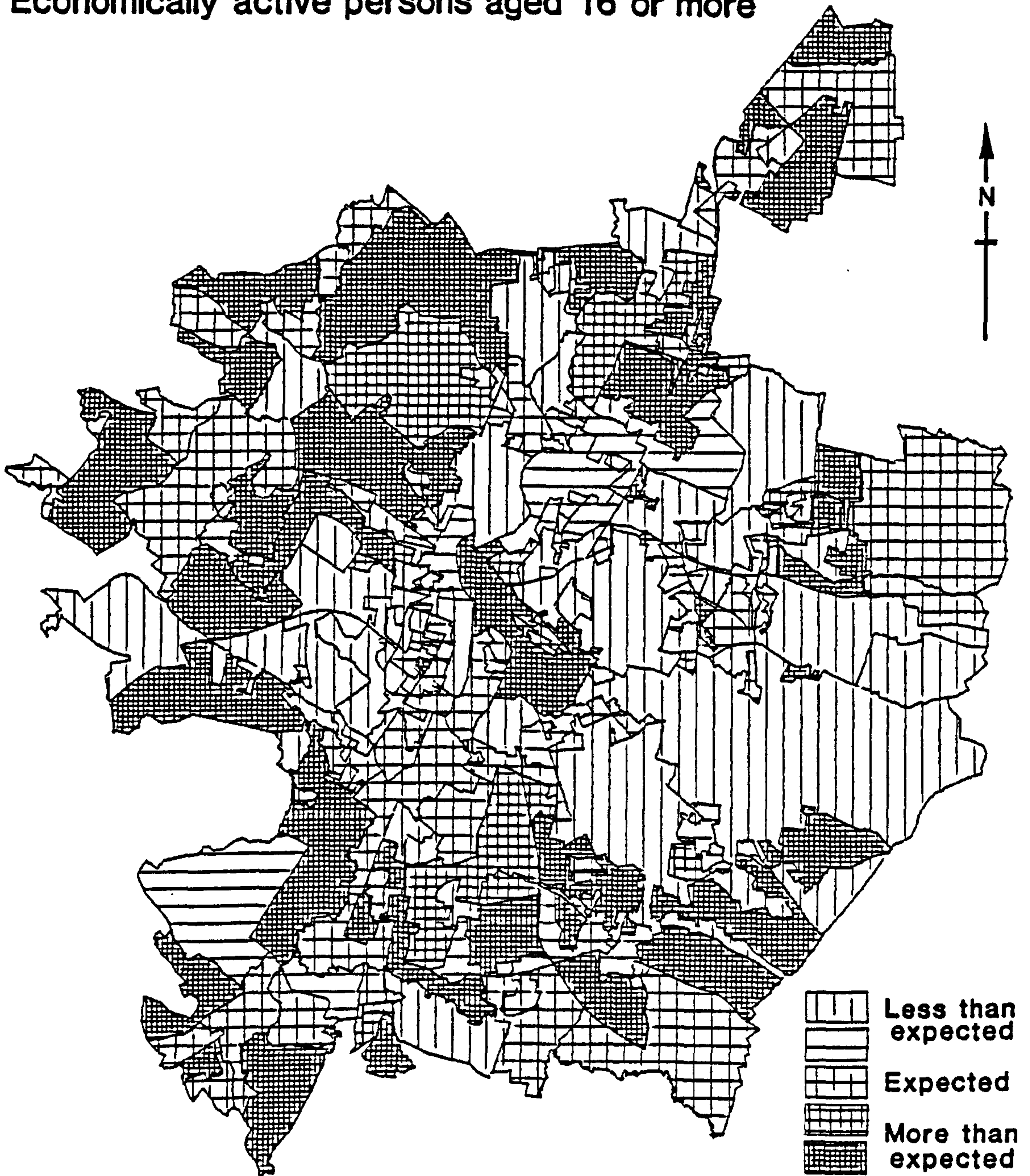
**Economically active persons aged 16 or more** (Figure 6.28)

The distribution of the economically active population coincides closely with that of the middle working age group (see Figure 6.5). Thus, above average proportions of economically active people are found in the city centre itself, and also in a semi-circular ring around the north, west and southern suburbs. To the east of the city and in a ring surrounding the city centre much lower proportions are found.

The main explanation of this pattern lies in the age

Figure 6.28

Economically active persons aged 16 or more



0 Kilometres 3

Mean value 64.77%

PLOTTED BY >> CHORO << 94/04/04.  
LEICESTER UNIVERSITY



distribution of the population, for example the inner suburbs have a predominantly older population. However, other factors may also contribute, such as the individual or household income and their needs and aspirations. These are likely to determine, for example, whether a housewife decides to make herself available for work and whether teenage and adult children continue in their education.

#### **Households with two or more economically active persons (Figure 6.29)**

About 45% of all households in Greater Leicester contain two economically active persons. Predictably, the distribution of these households coincides in general with that of the economically active population as shown on the last map (Figure 6.28), an exception is the city centre which has fewer households of this nature than expected. This is probably due to the dominance of single person households in the central area (see Figure 6.11), obviously having at most one economically active person.

The number of economically active persons in a household depends on several factors. For example the age structure of the household; the expectations of household members; family aspirations; and financial need. These factors influence both willingness and ability to look for work.

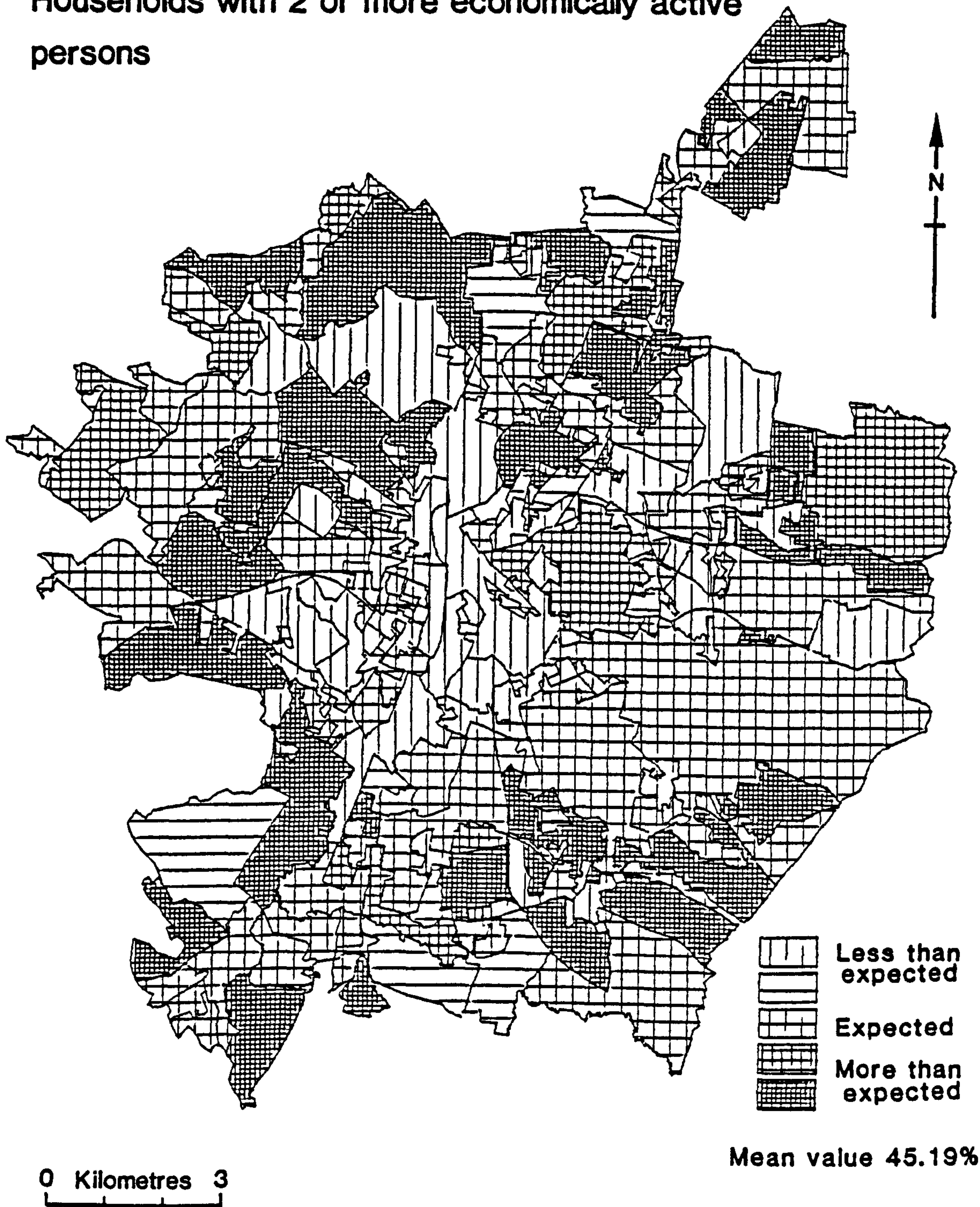
#### **Economically active persons seeking work (Figure 6.30)**

Overall, just over 9.5% of the economically active population are unemployed and seeking work, it is of major concern that these people should be so areally concentrated. In the outer suburbs there are almost no areas in which an 'expected' value is achieved, never mind exceeded, but within the city there are large zones of areas in the highest category. In particular the ethnic minority areas of Highfields, Belgrave and Spinney Hills, the older council estates of Braunstone, New Parks and Eyres Monseil,



Figure 6.29

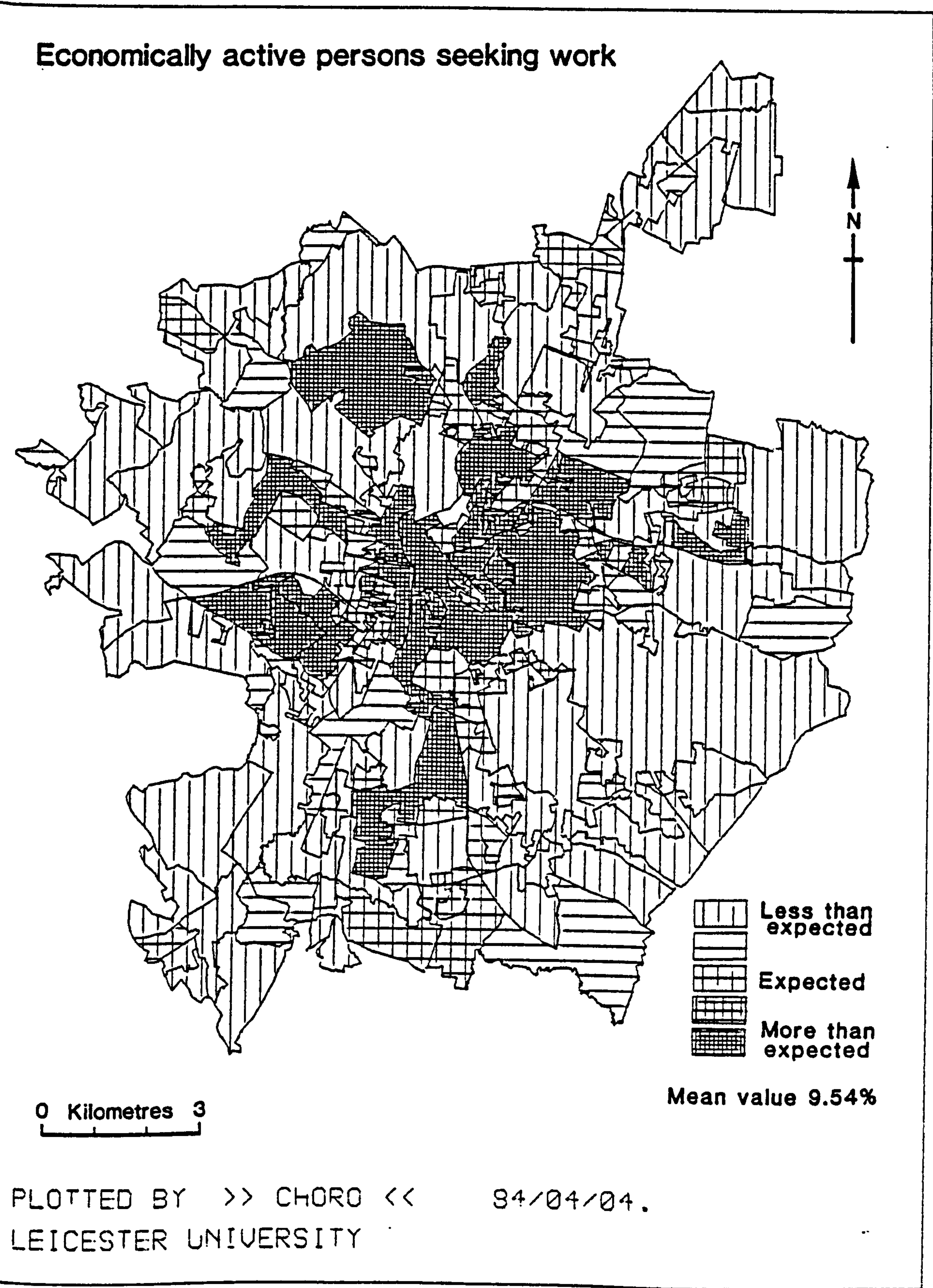
Households with 2 or more economically active persons



PLOTTED BY >> CHORO << 84/04/05.  
LEICESTER UNIVERSITY



Figure 6.30



and the inner city are badly affected. In part of Highfields 37.7% of economically active persons are seeking work.

There is no certain explanation to this pattern, other than the likelihood that people in the manual occupations are worst hit by unemployment during a recession. In Leicester the decline in the traditional manufacturing industries of footwear, textiles, and engineering is responsible.

#### **Unemployed people aged 16 to 24 (Figure 6.31)**

Unemployment among young people runs at 16.44%, significantly more than the average for the economically active population as a whole. There is extreme variation in this figure, from 0% to 100%, but this is due primarily to the problems associated with the use of ratio values for small populations.

As with unemployment in general, the areas having worst unemployment rates among the young include the older council estates, the ethnic minority areas and the city centre. A lack of job opportunities in general, and diminishing expectations in these areas in particular, are probably responsible.

In some respects unemployment among young people is more serious a problem than that among the overall population. Physically it threatens the fabric of the urban environment, socially it represents both current and future disruption of the community and in health terms it may be responsible for both physical and mental deterioration. Where unemployment is spatially concentrated these problems exacerbate each other and accelerate the downwards spiral.

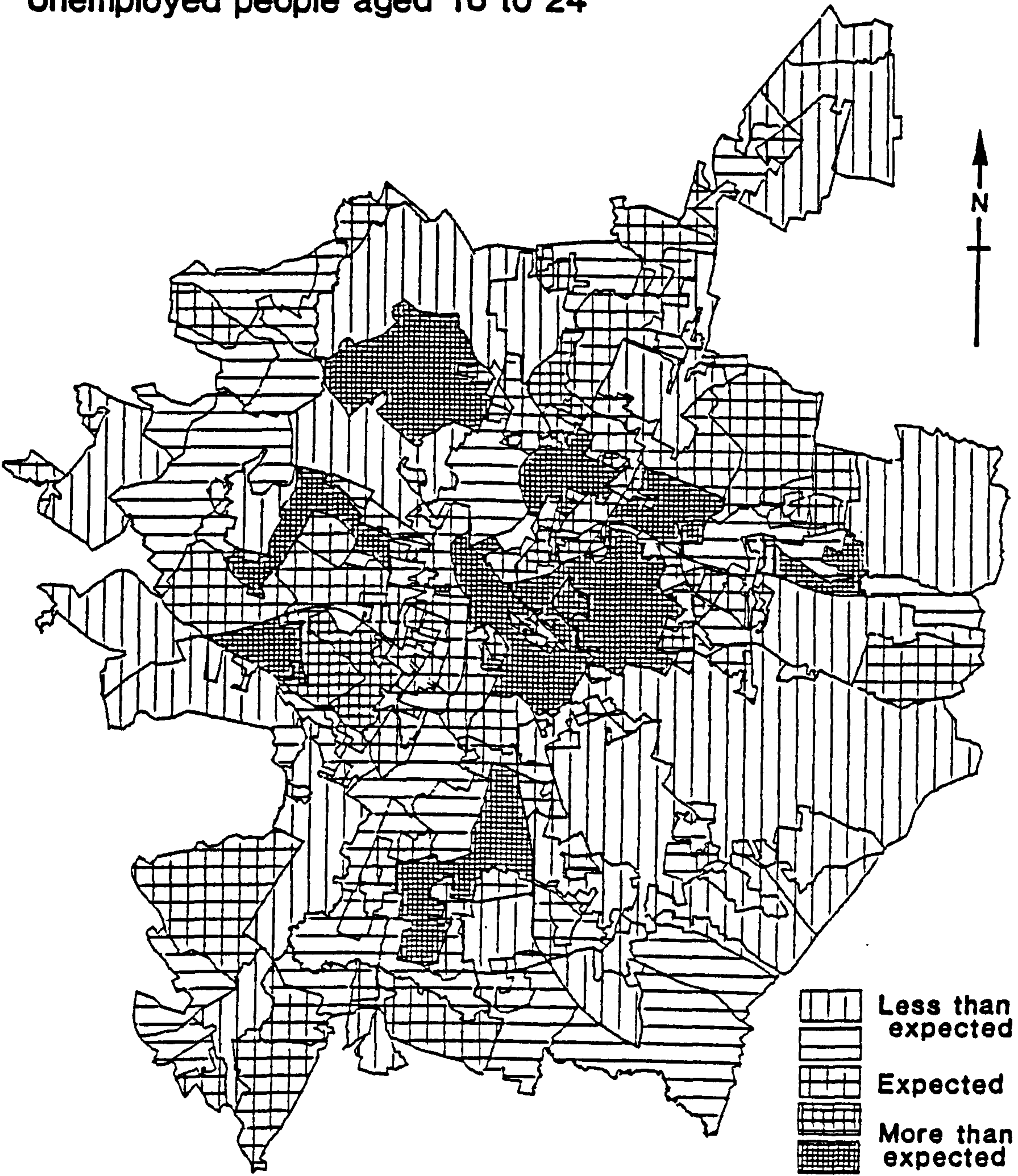
#### **Students (Figure 6.32)**

This category includes all residents aged 16 or more who are still in full-time education, just over 4.5% of this age group are so employed. Two factors contribute to the distribution of



Figure 6.31

Unemployed people aged 16 to 24

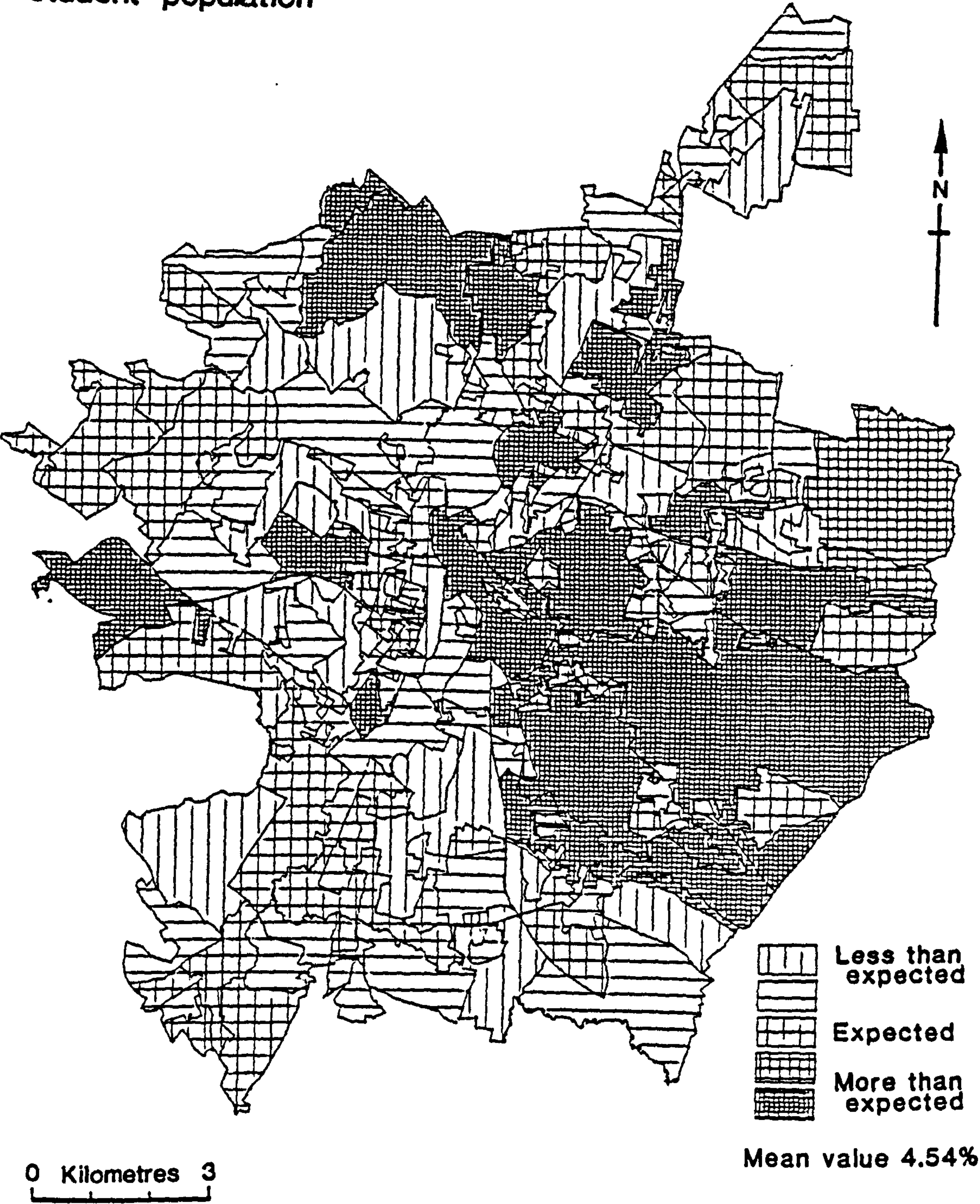


PLOTTED BY >> CHORO << 84/04/05.  
LEICESTER UNIVERSITY



Figure 6.32

Student population



PLOTTED BY >> CHORO << 84/04/04.  
LEICESTER UNIVERSITY

students in Greater Leicester. First, the distribution somewhat resembles that of the higher socio-economic groups. This supports the normally held view that the majority of students in higher education come from middle and high income families. For the purposes of the 1981 census most students were enumerated at their home address.

Secondly, there are above average proportions of students in the Highfields and Belgrave areas. This is due in part to the availability of privately rented accommodation for Leicester students especially those making their permanent home in the city, but more significantly, to the aspirations of the ethnic community. Having received comparatively little education themselves, at least within the British system, Asian parents are keen for their children to take advantage of the education facilities in Leicester. Whether the prospect of unemployment deters potential school-leavers among the ethnic community is uncertain.

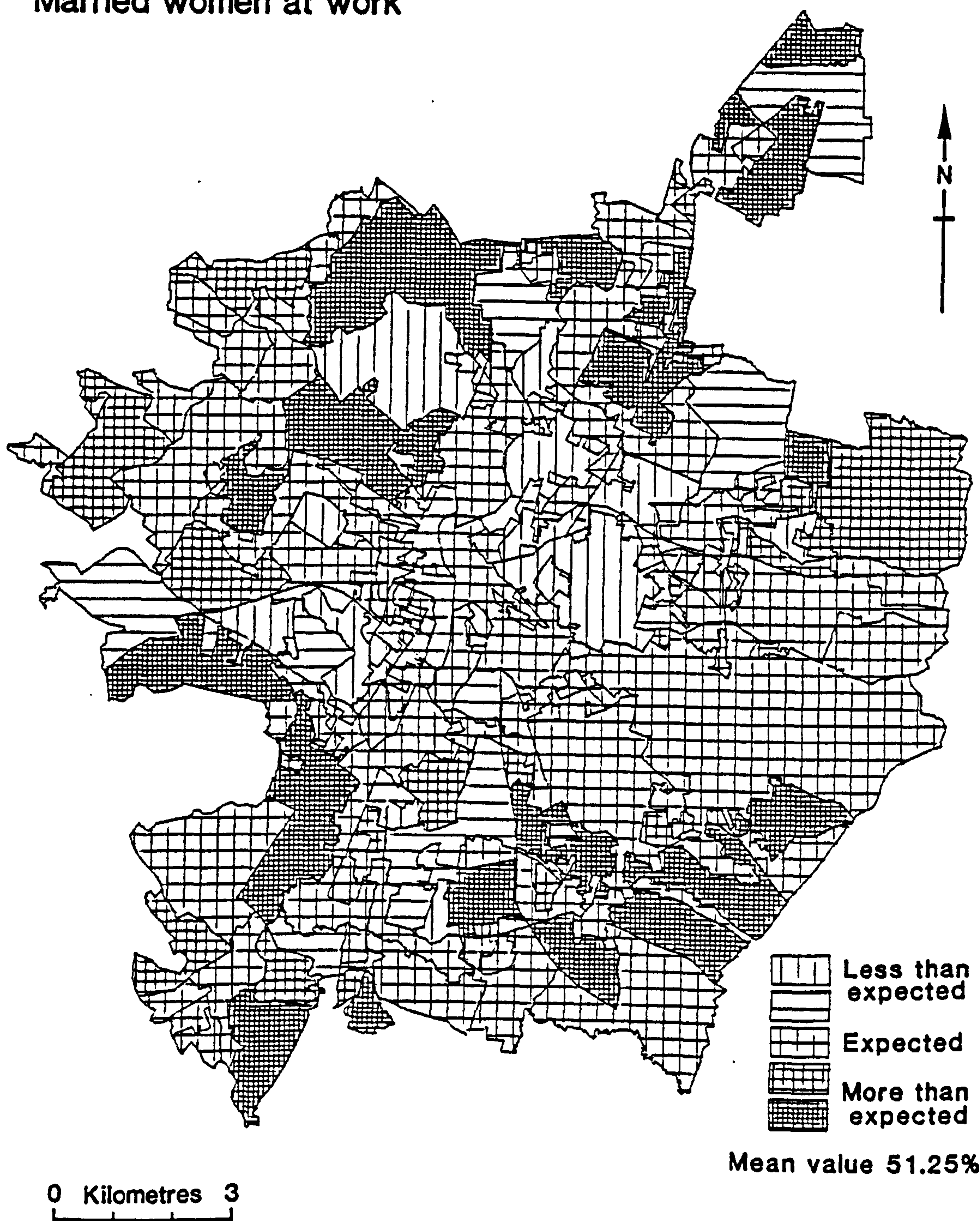
#### **Married women at work (Figure 6.33)**

Just over half of the married women in the study area are in either full- or part-time employment. The distribution of these is similar, but not identical, to that of households with two or more economically active people (see Figure 6.20) with middle class suburbia having the highest concentration. Most explanations are common to both. The discrepancies which may be seen can be accounted for by the choice of denominator, which gives the city centre for example, a higher proportion of married women at work than households with two economically active people. Alternatively, an explanation may lie in the household structure; where children rather than mothers go out to work there will be more households with several economically active people than there will be working married women. New Parks and Highfields provide



Figure 6.33

Married women at work



PLOTTED BY >> CHORO << 84/04/05.  
LEICESTER UNIVERSITY

examples of these.

#### **Self-employed persons (Figure 6.34)**

Although the term 'self-employed' can cover a wide range of professional as well as trades people, in Greater Leicester the distribution of the self-employed is most similar to that of professional and managerial headed households (see Figure 6.22). That is, above average proportions are found in the high status south-east sector and in isolated areas around the outskirts of the study area. There are generally less than average proportions among the council estates.

To explain this distribution requires certain assumptions regarding the motives for self-employment; specifically, is it assumed that an important motive is self-advancement, usually of a financial nature. Concomitant with this however, is social advancement, reflected by a person's place of residence. The distribution of the self-employed therefore depends on their housing aspirations, subject of course to cost and accessibility constraints.

#### **6.3.6 Other variables**

##### **One year migrants (Figure 6.35)**

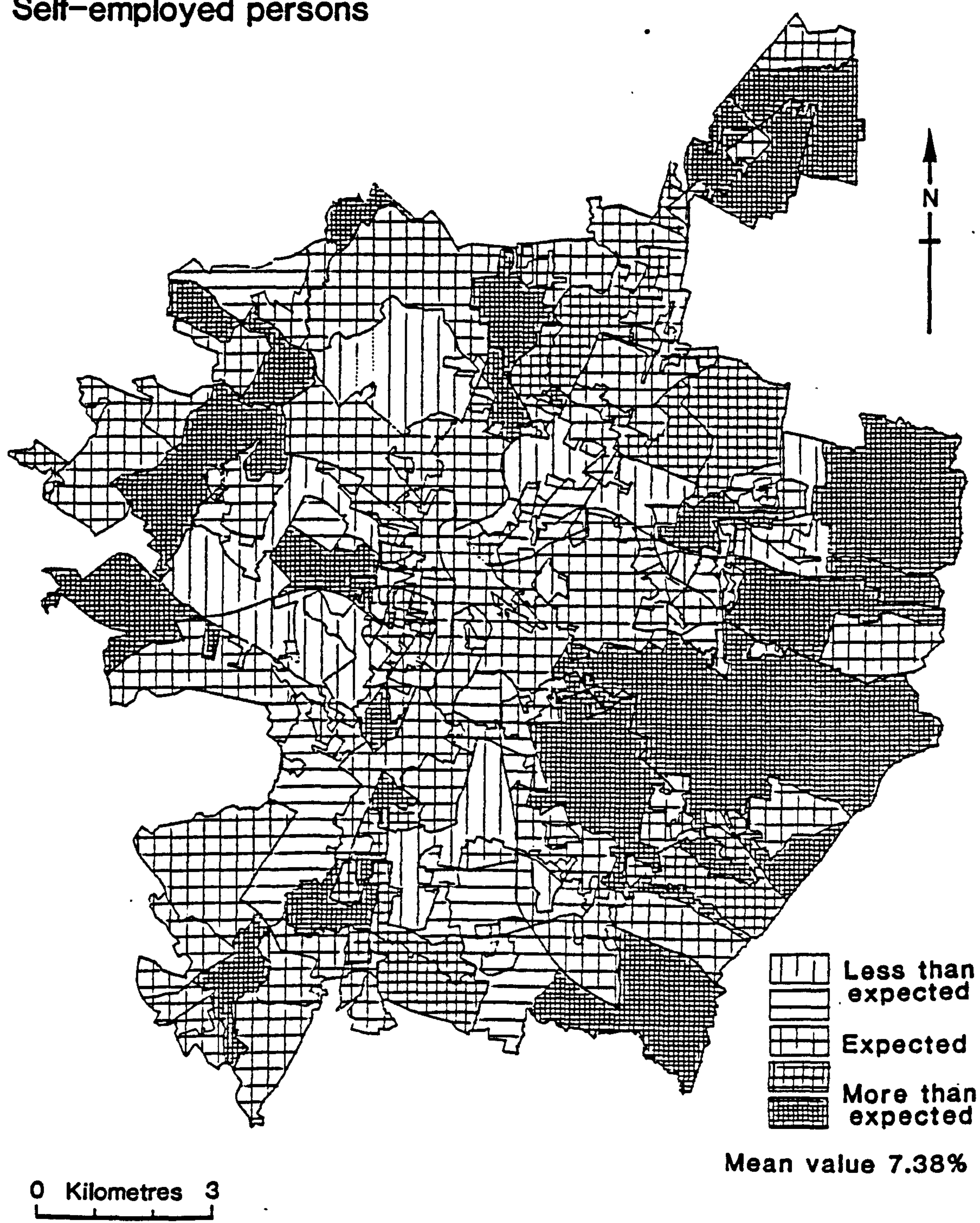
'One year migrants' include all residents aged over one year who were living at a different address one year prior to the census; they account for just over 9% of all residents in Greater Leicester.

The propensity to move is often associated with life cycle stage and changing residential requirements; on marriage, on the birth of a child, when the family leaves home and at retirement age, people are increasingly likely to seek a new home. In Greater Leicester there also appears to be a relationship with tenure. The highest proportions of one year migrants are found in the



Figure 6.34

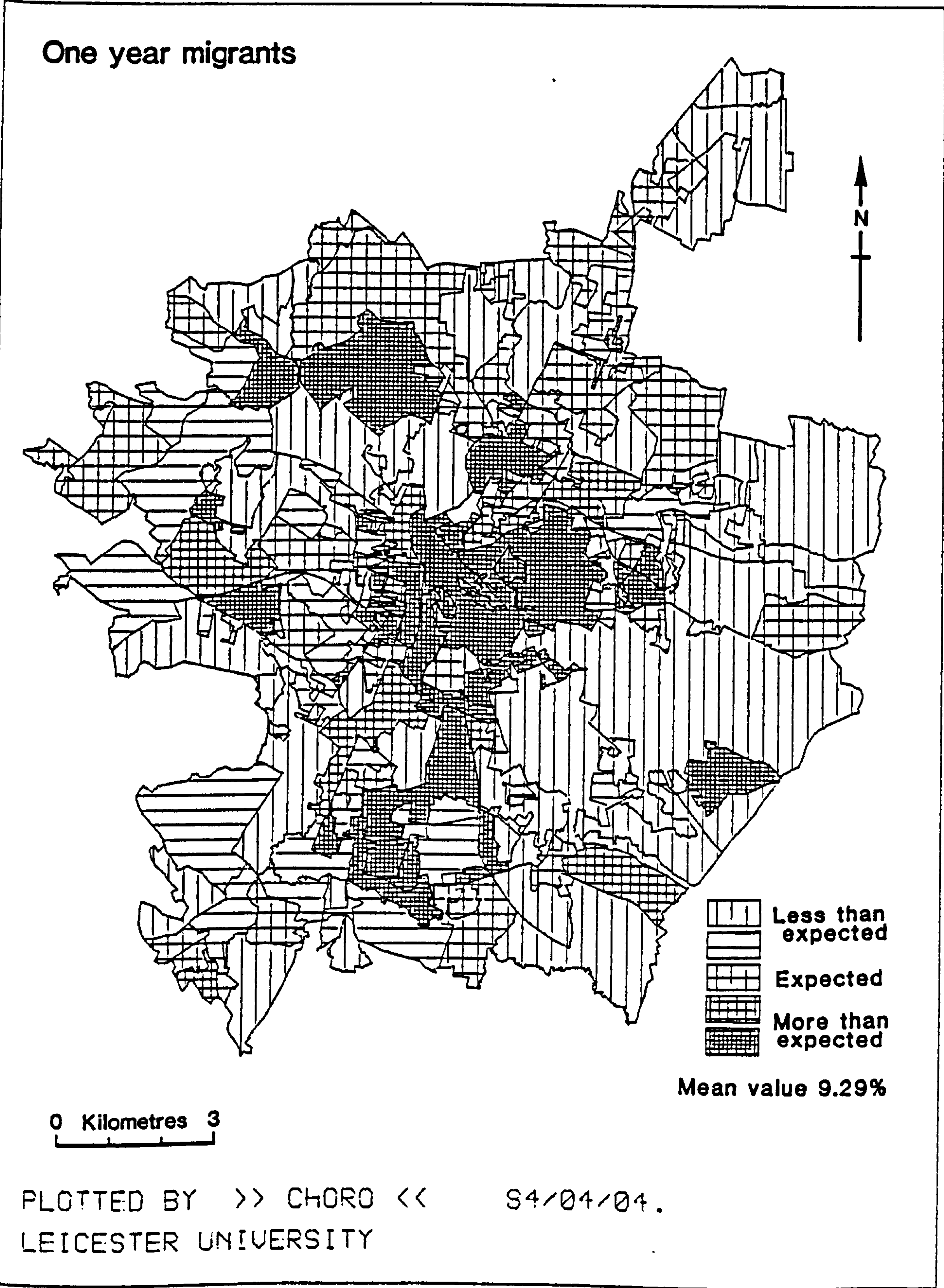
Self-employed persons



PLOTTED BY >> CHORO << 84/05/14.  
LEICESTER UNIVERSITY



Figure 6.35



nineteenth century city within the privately rented housing market. Above average proportions of one year migrants are found near the city centre between West End and North Evington and between Belgrave and Eyres Monsell. Within these areas the movement of young working people is likely to be a contributing factor, as is movement among the ethnic community as they consolidate their position in the housing market. To the north the high proportion of one year migrants in Beaumont Leys is explained by the amount of new building in that area.

#### **Households without a car (Figure 6.36)**

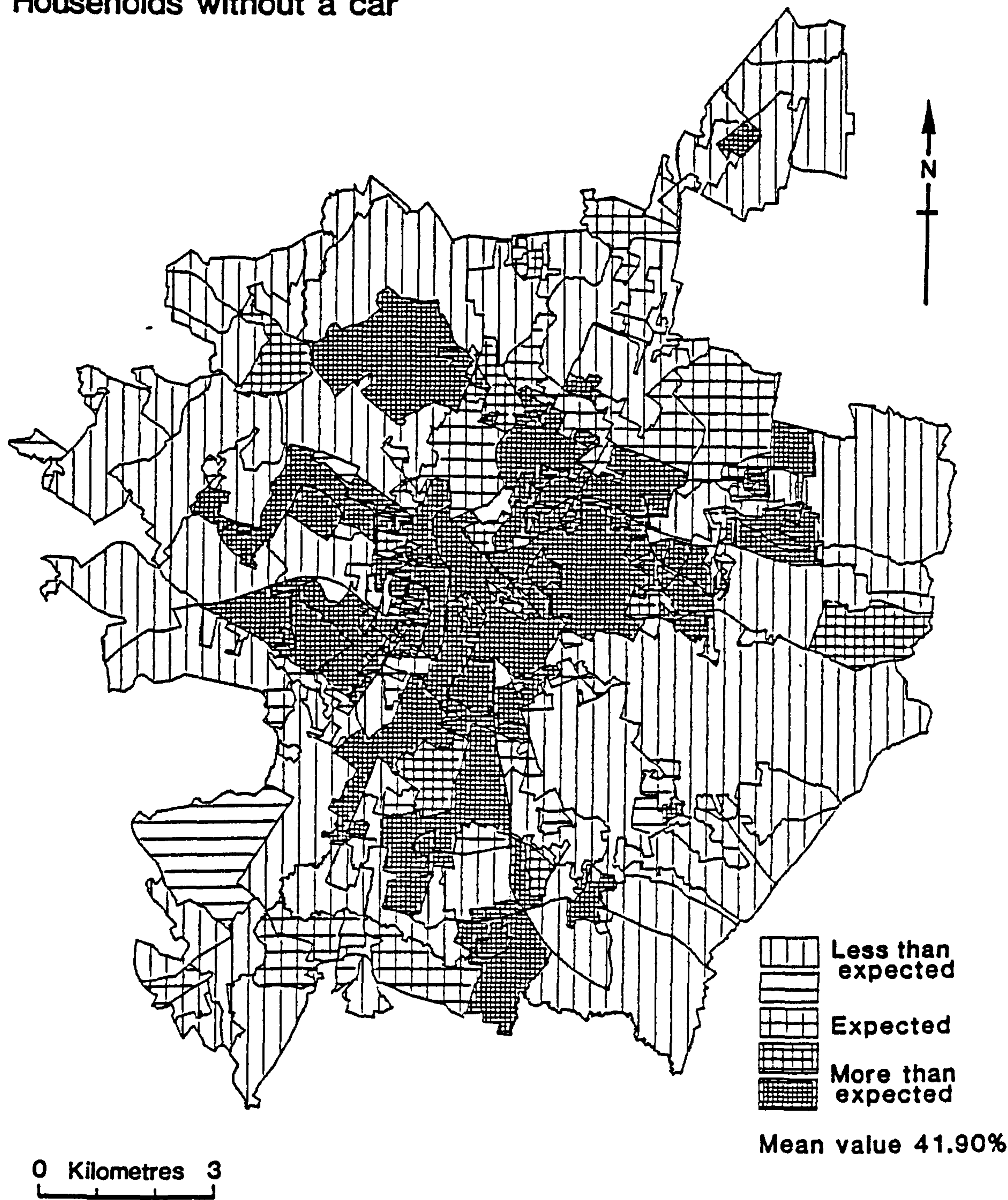
Despite generally increased levels of car ownership over recent years some 42% of households in Greater Leicester are still without. This figure varies enormously, from 2.8% in Anstey to 92.6% in part of Highfields. Three explanations are offered.

First, car ownership depends on the need for private transport, secondly, levels of car ownership are determined by wealth, and finally, car ownership is related to age. Thus suburban dwellers have the greatest need for private transport and in the case of the high status areas the better ability to pay for it. The elderly and those living in the less well-off areas have neither the ability nor the need for a car.



Figure 6.36

Households without a car



PLOTTED BY >> CHORO << 84/04/05,  
LEICESTER UNIVERSITY



## Chapter 7. Perinatal Mortality in Greater Leicester

### 7.1 Introduction

The previous two chapters have been concerned with the delimitation and description of new small areas within Greater Leicester. Each NSA was created from one or more EDs on the basis of their internal socio-economic and demographic structure. In particular, attention was focussed on measures of disadvantage or deprivation, such as unemployment, overcrowding and single parenthood. Certain parts of Leicester were found to be disadvantaged on the basis of several variables, these included the inner city, some of the older industrial village centres and the older council estates.

Perinatal mortality is itself frequently considered to be a measure of deprivation, it is used on an international scale as an indicator of social and economic well-being (see Newsholme, 1910 and Stockwell and Adamchak, 1978) and at a sub-national level as a measure of the efficiency of the obstetric services (MacVicar et al., 1977). At the urban scale lack of suitable data normally precludes the use of the perinatal mortality rate for identifying disadvantage.

The purpose of this chapter is to explore the perinatal problem in Leicester at both the individual and aggregate scales. Section 7.2 addresses the first of these by describing the survey data set at the simplest level, examining variations between cases and controls in aspects of health. Although crude, this enables comparison of the problem in Leicester with findings from other studies (see Chapter 3) and it highlights associations worthy of more detailed examination.

Section 7.3 introduces a second data set, the Leicestershire births data. As previously explained, the main purpose of these data is to standardise for the mortality data, but first each of

the variables is described. In addition to increasing our knowledge of birth characteristics in Greater Leicester this also enables some assessment of the mortality survey's control population.

For each of these two sections the method of analysis is limited to the use of simple descriptive statistics. Three subprograms of the "Statistical Package for the Social Sciences" (SPSS) are used: CONDESCRIPTIVE, FREQUENCIES and CROSSTABS. The first of these calculates mean values, standard deviations, maximum and minimum values and so forth; the second produces frequency tables for categorical type variables; and the third enables the significance of potential associations to be tested by means of the chi-square statistic.

In Section 7.4 both mortalities and births are allocated to EDs and thence to NSAs and clusters as identified in Chapter 5. The objectives of this are to standardise the mortality data in a spatial sense and to compare the distribution of perinatal mortalities with that of socio-economic disadvantage. This is done both statistically and cartographically.

The chapter concludes with an interpretation of maps showing the distribution of mothers exhibiting the 'risk' characteristics identified in Section 7.2. These are compared with both the overall patterns of mortalities and births as well as the distributions of the census indicators.

## **7.2 The Perinatal Mortality Survey - description of the data set**

Chapter 3 concluded with a list of variables (Table 3.3) which have been found in previous studies to be associated with perinatal mortality. In this section a selection of these and other variables will be considered in turn, comparing the Greater Leicester experience with previous findings.

### 7.2.1 The mother's physical characteristics and well-being

Five of the six 'physical' characteristics listed in Table 3.3 are pertinent to the present study; namely, maternal age and height, nutritional status, alcohol consumption and smoking habits. The sixth, the mother's year of birth, is not examined here. Although Baird (1976) for example, found that mothers born during the 1928-32 depression suffered an above average stillbirth rate, no comparative period of economic hardship has since occurred.

Many authors, for example Morris and Heady (1955), Elwood, Mackenzie and Cran (1974) and Peckham, Ross and Farmer (1982), agree that maternal age is an important explanatory variable. Specifically, from about 30 years upwards the risk to the infant increases with maternal age. Surprisingly, this result is not supported by the present study (see Table 7.1). The age distributions of case and control mothers are similar, as are the mean ages of 25.7 and 25.6 years respectively and their standard deviations. Only a slight excess of case mothers over controls is found in the higher (30 years and over) and lowest (16 years or less) age groups, neither being significant under the chi-square test.

Maternal height or physique is considered by some researchers (such as Butler and Bonham (1963), Baird and Thomson (1969)) to be an indicator of past nutritional status, general health and well-being. They have found that as maternal height increases, the likelihood of a perinatal death decreases. In the present survey control mothers are on average 1.0 cm. taller than case mothers and there is a significant difference in the frequency distribution of the two groups (see Table 7.2).

Nutritional status may be directly measured by examining



Table 7.1

Maternal Age	Cases		Controls	
	N	%	N	%
16 or less	11	1.6	9	1.3
17 to 19	85	12.4	95	13.9
20 to 24	213	31.1	201	29.4
25 to 29	219	32.0	235	34.4
30 to 34	104	15.2	96	14.0
35 or more	53	7.7	48	7.0
	<hr/>	<hr/>	<hr/>	<hr/>
	685	100.0	684	100.0

$\chi^2 = 2.23$  with 5 degrees of freedom  
Not significant

Table 7.2

Maternal height (cms)	Cases		Controls	
	N	%	N	%
150 or less	85	12.6	73	10.7
151-155	142	20.9	113	16.6
156-160	200	29.5	189	27.7
161-165	151	22.3	191	28.0
166-170	76	11.2	86	12.6
171-190	24	3.5	30	4.4
	<u>678</u>	<u>100.0</u>	<u>682</u>	<u>100.0</u>

$\chi^2 = 10.47$  with 5 degrees of freedom  
significance <0.1

Table 7.3

Parity	Cases		Controls	
	N	%	N	%
0	369	53.9	334	48.8
1	147	21.5	210	30.7
2	92	13.4	77	11.3
3	38	5.5	34	5.0
4	18	2.6	16	2.3
5	13	1.9	8	1.2
6 and over	8	1.2	5	0.7
	<u>685</u>	<u>100.0</u>	<u>684</u>	<u>100.0</u>

$\chi^2 = 16.41$  with 6 degrees of freedom  
significance <0.05

food and drink consumption. The most recent version of the survey questionnaire takes details of the mother's intake of milk, coffee, meat, fish and various 'ethnic' foods (see question 46 of the questionnaire in Appendix 3). Unfortunately earlier respondents were not asked about their eating habits so overall enumerated rates are low and statistically unreliable.

A similar constraint applies to the measurement of alcohol consumption. Currently there are several questions regarding alcohol intake; including whether the mother normally drinks alcohol, whether she did so during pregnancy, and if so, how much, of which type of drink, and during which months? (see question 46). In total, 103 case mothers and 102 controls gave details of their drinking habits; of these, more control mothers than cases stated that they drank alcohol both before and during pregnancy, but the difference is not statistically significant. There is therefore no evidence either to support the claim of Peckham et al. (1982) that alcohol consumption might increase the risk of spontaneous abortion, or to suggest that perinatal outcome is so affected.

The final factor to affect the mother's, and hence the baby's, physical well-being is cigarette smoking. Previous work (Butler, Alberman and Peel 1969; Llewellyn-Jones, 1974) has found that smoking affects fetal growth, birthweight and the length of the gestation period, resulting in increases in both stillbirths and neonatal deaths. As with alcohol consumption the survey differentiates between usual smoking habits and smoking during the different stages of the pregnancy. Overall, about 44% of mothers who answered these questions had ever smoked and 35% had continued smoking during pregnancy. Differences between case and control groups were minimal.

In summary, the Leicester survey shows little correspondence



with previous findings regarding the mother's physical characteristics. Only one variable, maternal height, shows a significant relationship with perinatal outcome, and this only at the 0.1 probability level. Maternal age and smoking habits exhibit the predicted relationship but without statistical significance, in the consumption of alcohol mothers in the control group surprisingly exceed cases.

### 7.2.2 Mother's obstetric history

Parity, birth interval and the outcome of previous pregnancies have frequently been found to be associated with subsequent perinatal death (Morris and Heady 1955; Butler and Bonham, 1963). Each of these factors have been included in the present survey, as well as additional information on gravidity and cause of previous infant deaths.

'Parity' refers to the number of viable (i.e. over 28 weeks gestation) infants previously born to the mother. All live and stillbirths are included, with multiple births increasing parity proportionately. The current pregnancy is excluded.

The literature suggests (Medical Statistics Division OPCS, 1978; Adelstein et al., 1980) that mothers of zero or very high parity are most at risk during their pregnancy. In the present study the proportion of case mothers in these high risk categories exceeds that of control mothers. Table 7.3 shows this to be significant at the 0.05 level.

Gravid state is a similar measure to parity but it also includes both previous abortions and the current pregnancy. The number of pregnancies is measured, not the number of infants. As with parity, the proportion of survey mothers with gravidity 1 or over 5 is higher for cases (43.9%) than for controls (41.3%), this too is statistically significant (see Table 7.4).

Table 7.4

Gravidity	Cases		Controls	
	N	%	N	%
1	265	38.8	256	37.5
2	173	25.3	219	32.0
3	121	17.7	104	15.2
4	61	8.9	48	7.0
5	29	4.2	31	4.5
6	15	2.2	17	2.5
7 and over	20	2.9	9	1.3
	<u>684</u>	<u>100.0</u>	<u>684</u>	<u>100.0</u>

$\chi^2 = 12.75$  with 6 degrees of freedom  
significance <0.05

Table 7.5

Previous pregnancy outcome	Cases		Controls	
	N	%	N	%
Dead	58	13.4	33	7.8
Alive	374	86.6	391	92.2
	<u>432</u>	<u>100.0</u>	<u>424</u>	<u>100.0</u>

$\chi^2 = 7.17$  with 1 degree of freedom  
significance <0.01

The length of time since the mother's last pregnancy, or the 'birth interval', has been shown to be associated with adverse outcome (Yerushalmy, 1956; Douglas 1950). Like maternal age, there appears to be an 'ideal' range for the birth interval, suggested by these authors as 18 to 35 months, above and below which the risks to the baby are greater. This however cannot be detected within the survey population since across yearly intervals cases and controls have essentially similar frequency distributions. In only two categories do case mothers noticeably exceed controls. The first is the one year birth interval which comprises 48 cases and 43 controls, the second is the interval of eleven or more years into which only 11 case mothers and 4 control mothers fall. Different explanations may account for these.

The greater number of case mothers in the one year category is explained by Chalmers' (1979b) theory of 'reproductive compensation'. He suggests that mothers suffering a pregnancy loss are more likely to swiftly fall pregnant again. If the baby from the second pregnancy subsequently dies this may not be due to the short birth interval, but rather to some other factor inherent to the mother and common to both pregnancies. Thus a short birth interval *per se* is not necessarily a risk factor. The same reasoning applies to high parity; the mother who successfully delivers several healthy babies need not be at any greater risk than her low parity counterpart, however, if she has a history of adverse pregnancy outcome and her high parity is a consequence of this, there is a legitimate risk.

The relatively high number of case mothers having a long birth interval may also have a multi-factorial explanation. Maternal age and possible deterioration in maternal health may account for an excess of cases in this category.

Medical textbooks, for example Llewellyn-Jones (1974),



suggest that the occurrence of one perinatal death increases the likelihood of another, the exact probability depending on the state and cause of death. For each mother the perinatal survey takes details of up to seven of her previous pregnancies; including for each of these the period of gestation, complications of the antenatal period and delivery, birthweight and infant outcome (see question 7). Here we are interested only in the latter, specifically, whether the baby survived or not. Table 7.5 includes all the previous pregnancies of all mothers in the survey, classified by outcome. It is shown that the proportion of previous pregnancies resulting in a death is consistently higher for mothers of cases than for those of controls. This pattern is accentuated when outcome is considered on the basis of alive and well/not alive and well (see Table 7.6). In this case the latter category includes babies that have either died or have lived but have suffered some type of abnormality or disease during gestation or birth (for example, prematurity, birth injury, malformation or accident). For this dichotomy the chi-square test indicates a highly significant association between previous and current pregnancy outcomes.

To summarise, of the four variables considered here, three have been found to have a statistically significant association with perinatal mortality, but for only one is this unqualified. Parity and gravidity are significantly associated with perinatal death according to the chi-square test, but it cannot be shown that there is a direct causal relationship. Rather, high parity may be the result of 'reproductive compensation' following pregnancy loss, the latter forming in reality the association with perinatal mortality. This hypothesis is supported by the strong relationship between previous pregnancy outcome and perinatal

Table 7.6

Previous pregnancy outcome	Cases		Controls	
	N	%	N	%
Alive and well	317	73.4	357	84.2
Other	115	26.6	67	15.8
	<u>432</u>	<u>100.0</u>	<u>424</u>	<u>100.0</u>

$\chi^2 = 47.7$  with 1 degree of freedom  
significance <0.001

Table 7.7

Mother's social class	Cases		Controls	
	N	%	N	%
1. Professional )	44	6.4	69	10.1
2. Intermediate )				
3. Skilled	182	26.7	185	27.0
4. Semi-skilled	142	20.8	105	15.4
5. Unskilled	27	4.0	18	2.6
6. Housewife	287	42.1	307	44.9
	<u>682</u>	<u>100.0</u>	<u>684</u>	<u>100.0</u>

$\chi^2 = 13.57$  with 4 degrees of freedom  
significance <0.01

mortality shown in the survey, particularly if all problem outcomes are examined.

### 7.2.3 Socio-economic characteristics

A wide range of variables have been suggested (see Table 3.3) in the search for an association between socio-economic status and perinatal outcome. Relationships between perinatal mortality and some of these, for example 'social class', are undisputed, (Lambert, 1976; Medical Statistics Division OPCS, 1978; Peckham et al., 1982 and others); but for some of the less easily quantified characteristics such as 'industrialisation', 'affluence', 'migrancy' and 'social disadvantage', authors can only hypothesise a 'multivariate' association (Thompson, 1968).

In the present study the socio-economic environment is considered to be a key factor affecting the social and spatial distribution of perinatal deaths. At the aggregate level the social structure of Greater Leicester has been fully documented (see Chapter 6), the perinatal survey complements this by permitting the investigation of the relationship between perinatal outcome and socio-economic characteristics at the individual level. The following variables are included in the data set

- mother's and father's social class and socio-economic group
- mother's marital status, whether she cohabits and the legitimacy of her child
- mother's country of birth, her religion and ethnic group
- employment during pregnancy.

As mentioned above, the relationship between social class and perinatal mortality is widely acknowledged. Specifically, people in the semi-skilled and unskilled occupations have higher perinatal death rates than their skilled and professional counterparts (Macfarlane, 1979a; Adelstein et al., 1980). Table



7.7 gives the breakdown into social classes of case and control mothers. It excludes mothers in the armed forces, students, and those for whom information is not available. There are significantly fewer case mothers than controls in the professional and intermediate classes and many more in the semi-skilled and unskilled groups. This supports previous findings.

Over one third of all mothers, however, gave their occupation as 'housewife' - not in itself a conventional social class. It does nevertheless have implications for perinatal outcome. In particular, since employment during pregnancy is thought to have a detrimental effect on birth outcome (McDowall and Goldblatt, 1981) and a 'housewife' is by definition not in employment, then a lower perinatal mortality rate for housewives is to be expected. The excess of controls, both absolutely and proportionately, over cases in the housewife category lends support to this hypothesis. A reduced perinatal mortality rate among housewives might, however, be due to factors other than employment. For example, many young women will only become a housewife when they have had their first child; having had a baby their parity increases and, with this, the risk of losing a subsequent baby decreases. Thus the effects of housewife status and parity become confused. Although the survey includes specific details of employment during pregnancy (see question 39) that part of the question which refers to the month of stopping work has only recently been introduced and as yet the results are inconclusive. It is not considered here.

To avoid the confusing influence of the 'housewife' category, the father's social class might be considered instead. This is not infrequent in social studies, in fact major data sources such as the census base the socio-economic and social

status of an entire household on the occupation of the household head. Allocating survey cases and controls to social classes according to the father's occupation shows the same excess of controls in the upper classes and cases in the lower classes, but in contrast with the findings from the mother's social class, the results are not significant under the chi-square test.

A more detailed picture of social status can be obtained by examining the socio-economic groups into which both parents fall. These are shown in Table 7.8. Among the mothers, the controls have a higher proportion in socio-economic groups 1,2,3,5 and 6, while cases better represent groups 4,7,8,9,10,11,12,13 and 77. With the exception of group 4 (professionals - employees) the former are the non-manual and the latter the manual categories.

Examination of the father's socio-economic groups reveals a more complex pattern. There are a higher percentage of controls than cases in groups 1,2,4,5,8,9,12 and 16, and a higher percentage of cases in the remainder. If it is possible to generalise, it may be said that the cases have greater representation in the lower status sectors of both white and blue collar groups i.e. 'junior non-manual', 'personal service workers', 'semi-skilled' and 'unskilled' workers. Among the professional, managerial and skilled workforce the controls are dominant.

To test the significance of these observations, in Table 7.9 the data have been aggregated into five socio-economic classes. For both the mothers' and the fathers' socio-economic class there is a significant difference between the case and control populations, the former being biased to the manual categories and the latter to the non-manual occupations.

Marital status and, in particular, legitimacy, have been found to be associated with perinatal mortality (Lambert, 1976;

Table 7.8

Socio-economic group	Cases		Controls		Cases		Controls	
	N	%	N	%	N	%	N	%
~~~~~								
1. Large employers/managers	-	-	2	0.3	5	0.7	7	1.0
2. Small employers/managers	5	0.7	11	1.6	52	7.6	60	8.8
3. Self employed professionals	-	-	1	0.1	7	1.0	3	0.4
4. Professionals - employees	7	1.0	2	0.3	22	3.2	37	5.4
5. Intermediate non-manual	43	6.3	55	8.1	36	5.3	42	6.1
6. Junior non-manual	90	13.2	118	17.3	55	8.0	52	7.6
7. Personal service workers	24	3.5	23	3.4	5	0.7	3	0.4
8. Foremen and supervisors - manual	7	1.0	1	0.1	20	2.9	33	4.8
9. Skilled manual workers	64	9.3	55	8.0	229	33.5	241	35.3
10. Semi-skilled manual workers	120	17.6	84	12.3	108	15.8	92	13.5
11. Unskilled manual workers	25	3.7	18	2.6	30	4.4	15	2.2
12. Own account workers	8	1.2	7	1.0	26	3.8	41	6.0
13. Farmers - employers and managers	1	0.1	-	-	-	-	-	-
14. Farmers - own account	-	-	-	-	-	-	-	-
15. Agricultural workers	-	-	-	-	4	0.6	1	0.1
16. Armed forces	-	-	-	-	3	0.4	4	0.6
17. Inadequately described + housewife + unemployed	287	41.9	305	44.6	39	5.7	28	4.1
77. Students	2	0.3	-	-	6	0.9	6	0.9
88. 'Not applicable'	1	0.1	2	0.3	37	5.4	19	2.8
99. 'Not available'	1	0.1	-	-	1	0.1	-	-
	<u>685</u>	<u>100.0</u>	<u>684</u>	<u>100.0</u>	<u>685</u>	<u>100.0</u>	<u>684</u>	<u>100.0</u>

Table 7.9

Socio-economic class	Mother				Father			
	Cases		Controls		Cases		Controls	
	N	%	N	%	N	%	N	%
1. Professional and managerial	13	3.3	16	4.2	86	14.4	107	17.1
2. Non-manual	133	33.8	173	45.9	91	15.2	94	15.0
3. Skilled manual	79	20.1	63	16.7	275	45.9	315	50.2
4. Semi-skilled	144	36.5	107	28.4	117	19.5	96	15.3
5. Unskilled	25	6.3	18	4.8	30	5.0	15	2.4
	394	100.0	377	100.0	599	100.0	627	100.0
Mothers' socio-economic class:		$\chi^2 = 13.57$ with 4 degrees of freedom significance <0.01						
Fathers' socio-economic class:		$\chi^2 = 11.48$ with 4 degrees of freedom significance <0.05						

Table 7.10

Legitimacy	Cases		Controls	
	N	%	N	%
Legitimate (Married + separated)	585	85.4	600	87.7
Illegitimate (Single + widowed + divorced)	100	14.6	84	12.3
	585	100.0	684	100.0

$\chi^2 = 1.58$ with 1 degree of freedom
significance <0.25

Adelstein et al., 1980; Peckham et al., 1982). Proportionately more perinatal deaths occur among illegitimate babies than among those born within marriage. The perinatal survey records both marital state of the mother and, if she is widowed, separated or divorced, whether she is cohabiting or not (see question 38). Differences between case and control groups in each of the marital states are minimal. Even if these categories are aggregated to a legitimate/illegitimate classification (see Table 7.10) the difference, although in an expected direction, is not statistically significant.

Cohabitation is not normally considered in studies of perinatal mortality, despite its potential use as an indicator of paternal support. Thus, the mother who claims to be cohabiting might reasonably be expected to receive both financial and moral support from the father for herself and her child. Of the unmarried mothers in the survey 46.4% of case mothers and 55% of control mothers stated that they were cohabiting. As with legitimacy, these figures differ in the expected direction, but are not significant under the chi-square test.

The importance of the Asian community to the social geography of Leicester was highlighted in Chapter 6. Some authors (for example Adelstein et al., (1980)) have found country of birth to be associated with perinatal outcome, that is to say, mothers born in the New Commonwealth or Pakistan have higher perinatal mortality rates. Within the current data set ethnicity is indicated by both country of birth and ethnic group.

Mothers are ascribed to an ethnic group 'using observation plus conversational probing' (see question 42), the five main ethnic groups are included in Table 7.11. Key features of this table are the excess of case mothers over controls in the 'Indian'

Table 7.11

Mother's ethnic group	Cases		Controls	
	N	%	N	%
U.K.	462	67.4	509	74.4
Eire	10	1.5	9	1.3
Indian	178	26.0	126	18.4
Pakistani/Bangladesh	11	1.6	11	1.6
West Indian	13	1.9	8	1.2
Other	11	1.6	21	3.1
	<u>685</u>	<u>100.0</u>	<u>684</u>	<u>100.0</u>

$\chi^2 = 15.5$ with 5 degrees of freedom
significance <0.01

Table 7.12

Mother's region of birth	Cases		Controls	
	N	%	N	%
Britain	459	67.0	507	74.1
Asia (India, Pakistan, Other Asia + Oceania)	91	13.3	63	9.2
New Commonwealth Africa and America (excluding U.S.A.)	110	16.1	90	13.2
Other (Eire, Old Commonwealth, Europe, elsewhere)	25	3.6	24	3.5
	<u>685</u>	<u>100.0</u>	<u>684</u>	<u>100.0</u>

$\chi^2 = 15.2$ with 3 degrees of freedom
significance <0.005

category, and the relative under-representation of cases in the 'U.K.' class. These differences are significant to the 0.01 level.

The mother's **country of birth** is ascertained by a direct question and is hence more reliably identified. Sixteen areas of birth are defined which are then aggregated into four classes corresponding to the major ethnic groups in Greater Leicester (see Table 7.12). Two points are noted. First, the number of British-born mothers may include a proportion of second generation immigrants; and secondly, the 'New Commonwealth' category may include some people of Asian descent born in Africa. There is strong evidence that the latter is true, given the distribution in Table 7.11. The chi-square statistic for Table 7.12 indicates a highly significant difference between the case and control groups, with a particularly large proportion of mortalities in the Asian category.

Ethnicity and religious preference are closely related. Although an association between religion and perinatal death has not previously been reported it may be hypothesised that such a relationship exists, explained, for example, by behavioural or dietary differences among the religious groups. Four religions are coded in the survey; Hindu, Moslem, Sikh and Christian (see Table 7.13). A statistically significant difference is observed between the distribution of cases and controls, with Hindus in particular, having disproportionate representation among the mortalities. The relatively low response should be noted, although assuming that the non-respondents are Christian mothers (this seems likely given the distribution across the ethnic groups in Table 7.11), this probably does not lessen the significance of the chi-square value.

The length of time for which the mother has been resident in the U.K. is a final indicator of ethnicity. It is suggested that **length of residence** is a crude measure of cultural assimilation,

Table 7.13

Mother's religion	Cases		Controls	
	N	%	N	%
Hindu	117	44.5	68	32.5
Moslem	32	12.2	40	19.1
Sikh	36	13.7	26	12.5
Christian	78	29.6	75	35.9
	<hr/>	<hr/>	<hr/>	<hr/>
	263	100.0	209	100.0

$\chi^2 = 9.5$ with 3 degrees of freedom
significance <0.05

Table 7.14

Infant birthweight (grammes)	Cases		Controls	
	N	%	N	%
200 or less	396	57.8	9	1.3
2001-2250	39	5.7	17	2.5
2251-2500	65	9.5	36	5.3
2501-3000	85	12.4	156	22.8
3001-3500	48	7.0	252	36.8
3501-4000	37	5.4	164	24.0
over 4000	15	2.2	50	7.3
	<hr/>	<hr/>	<hr/>	<hr/>
	685	100.0	684	100.0

$\chi^2 = 655.1$ with 6 degrees of freedom
significance <0.001

mothers having spent many years in this country being more closely absorbed into British society. Furthermore, the mother who has been here for many years has benefitted from our welfare state, receiving a comparatively high level of health care, education and financial reward. It might therefore be expected that the longer the mother has been in the U.K., the lower the likelihood of a perinatal death. The data however do not support this hypothesis since in only one of five categories, that of five to nine years residence in the U.K., do cases markedly exceed controls.

Of the socio-economic variables considered in this section, six have been found to be associated with perinatal mortality, namely, the mother's social class, the socio-economic class of both parents, the mother's ethnic group, her country of birth and her religion. Of the remainder, three variables showed the expected difference between cases and controls, even if the degree of variation was not significant under the chi-square test. These are the father's social class, legitimacy and cohabitation. On the whole, these results support the findings of other authors (see Spicer and Lipworth (1966), Lambert (1976), Medical Statistics Division, OPCS (1978), Macfarlane (1979a)).

7.2.4 Infant characteristics

Epidemiologists have invested much effort in seeking an explanation of perinatal death in the various characteristics of the infant. Birthweight, gestational age, length and type of labour, mode of delivery, multiplicity, infant sex and time or season of birth have all been suggested to be associated with pregnancy loss. These will be considered in this section.

Birthweight is almost universally agreed to be singly the most significant factor in explaining perinatal death (eg. Butler and Bonham, 1963; Baird, 1976; Peckham et al., 1982). In the

current survey both the distribution of cases and controls across birthweight categories, (Table 7.14), and the extreme difference in the mean values for the two groups (1837g for cases and 3237g for controls) point to a high level of association between low birthweight and perinatal mortality.

Birthweight, broadly speaking, depends on the duration of the pregnancy at delivery. Thus, babies born after the normal gestational period, that is, between 37 and 42 completed weeks following the last menstrual period (L.M.P.), are likely to weigh somewhere in the range of 2500g to 4000g. Babies born prematurely will not have had the opportunity to develop to their full potential and will therefore be most at risk. In addition, babies that have developed slowly in utero, known as 'small-for-dates' babies (Wallis and Harvey (1980)), will be similarly at risk. Figure 7.1 shows the variation between cases and controls in gestational age. Clearly the proportion of cases born 'pre-term' greatly exceeds the proportion of controls, while the converse is true for 'post term' births. The latter would seem to contradict previous evidence that the risk of perinatal death increases with a gestational age of greater than 40 weeks (Butler and Bonham, 1963). In Table 7.15 gestational age is classified into the three categories: pre-term, term and post-term. The pre-term class is dominated by cases, the others by controls. With a chi-square value of 512.5 the association between gestational age and perinatal death is highly significant.

Attributes of the labour might play an important role in the outcome of the pregnancy, for example Butler and Bonham (1963) have suggested that the length of time spent in labour influences perinatal outcome; or specifically, if labour continues for over 24 hours the risk to the baby increases. A total of fourteen women in the survey, comprising nine case mothers and five controls were

Table 7.15

Gestational age	Cases		Controls	
	N	%	N	%
Pre-term (36 weeks or less)	429	62.6	33	4.8
Term (37 to 41 weeks)	236	34.5	584	85.5
Post term (42 weeks or more)	20	2.9	67	9.7
	<u>685</u>	<u>100.0</u>	<u>684</u>	<u>100.0</u>

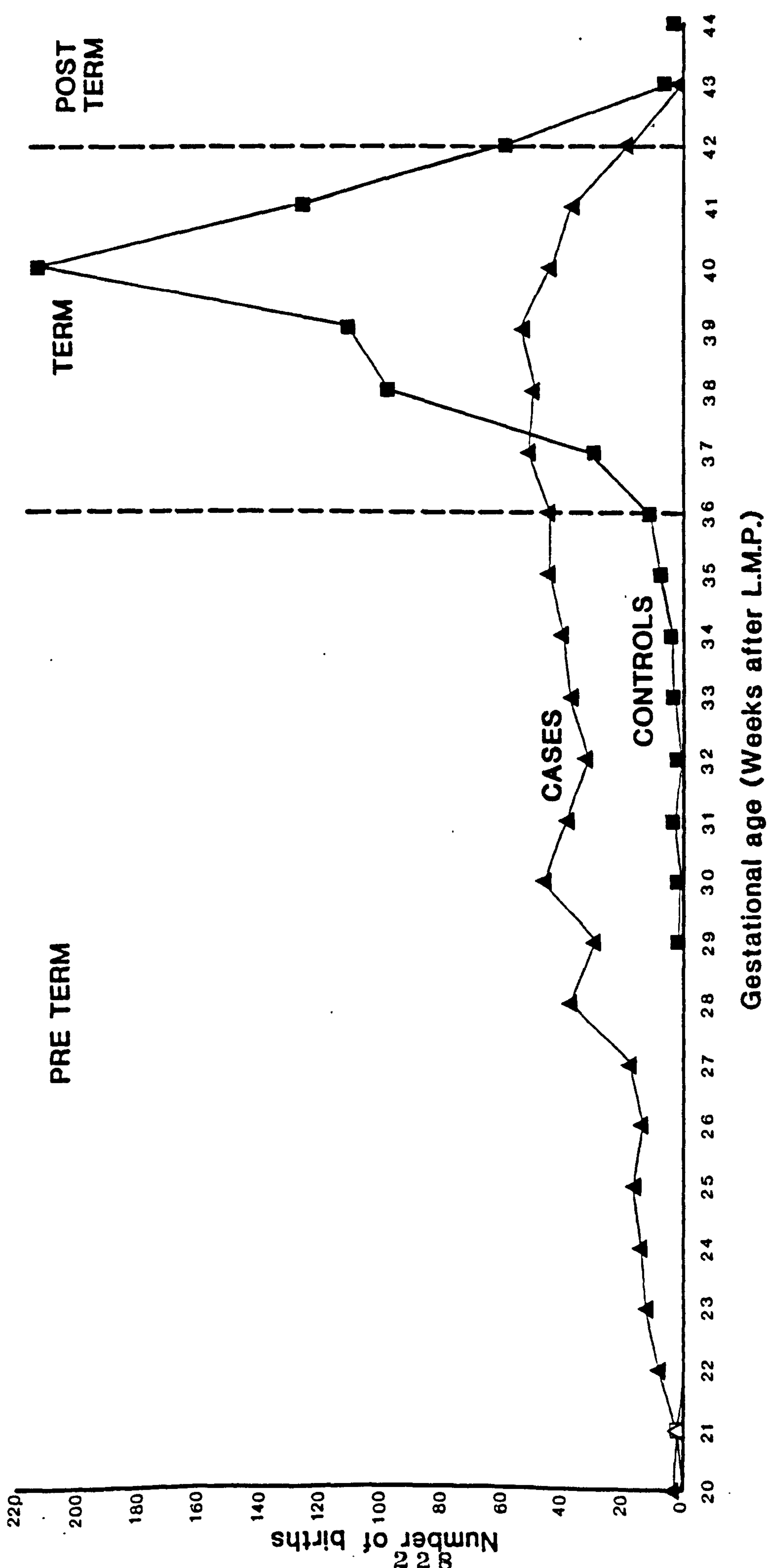
$\chi^2 = 512.5$ with 2 degrees of freedom
significance <0.001

Table 7.16

Type of labour	Cases		Controls	
	N	%	N	%
Did not labour	26	3.9	22	3.2
Spontaneous	262	38.2	156	22.8
Surgical induction	88	12.8	139	20.3
Medical induction - prostaglandins	91	13.4	17	2.5
Medical induction - oxytocics	13	1.9	11	1.6
Augmented labour - ARM	75	10.9	212	31.0
Augmented labour - oxytocics	62	9.1	55	8.1
Augmented labour- ARM + oxytocics	<u>67</u>	<u>9.8</u>	<u>72</u>	<u>10.5</u>
	<u>684</u>	<u>100.0</u>	<u>684</u>	<u>100.0</u>

$\chi^2 = 155.5$ with 7 degrees of freedom
significance <0.001

Figure 7.1 Gestational age - Cases and controls



in labour for more than 24 hours. The survey's maximum value, over 63 hours in labour, was achieved by a control mother. Even if the mothers are aggregated to four classes, less than 2 hours, 2 to 12 hours, 12 to 24 hours and over 24 hours, there is no significant difference between case and control groups.

The **type of labour**, whether spontaneous or induced, may also bear some relationship with perinatal mortality, if only because in some cases induction is performed in response to a problem condition. Table 7.16 shows the types of labour recorded in the survey. Interestingly, less than half of the mothers in the survey went through labour spontaneously, of the control mothers this figure was less than one quarter. Among the cases, medical induction of one type or another was used on a significantly large number of women, 104 compared with 28 controls; while among the controls augmented labour was especially common, 339 control mothers to 204 cases. The chi-square test suggests a significant difference between the two groups.

The **mode of delivery** may also be examined. Approximately three-quarters of all mothers had a normal, vertex or breech, delivery. Slightly more controls than cases were delivered with forceps, and a few more cases than controls had caesarian births. Nine of the cases were delivered by breech extraction following death. Overall, differences between the two groups are relatively minor and are not statistically significant.

A more important pattern emerges when the **presentation of the fetus** is examined. Table 7.17 highlights the extremely large and statistically significant difference between cases and controls in the number and proportion of breech presentations. These are known to be associated with births early in the pregnancy (Llewellyn-Jones, (1974) suggests that at the 20th week

Table 7.17

Presentation at birth	Cases		Controls	
	N	%	N	%
Cephalic	486	71.8	657	96.1
Breech	146	21.6	22	3.2
Other	44	6.6	5	0.7
	<u>676</u>	<u>100.0</u>	<u>684</u>	<u>100.0</u>

$$\chi^2 = 148.1 \text{ with 2 degrees of freedom}$$

significance <0.001

Table 7.18

Multiplicity	Cases		Controls	
	N	%	N	%
Singleton birth	615	89.8	674	98.5
Part of a multiple birth	70	10.2	10	1.5
	<u>685</u>	<u>100.0</u>	<u>684</u>	<u>100.0</u>

$$\chi^2 = 47.7 \text{ with 1 degree of freedom}$$

significance <0.001

Table 7.19

Attendances for antenatal care after 20 weeks gestation	Cases		Controls	
	N	%	N	%
0	21	3.1	0	0
1- 5	314	46.9	46	6.8
6-10	250	37.3	319	46.8
11-15	76	11.4	301	44.2
16-30	9	1.3	15	2.2
	<u>670</u>	<u>100.0</u>	<u>681</u>	<u>100.0</u>

$$\chi^2 = 364.6 \text{ with 4 degrees of freedom}$$

significance <0.001

of pregnancy more than 40% of babies present by the breech) and are a cause of risk to both the mother, should operative intervention be required, and to the fetus. Butler and Bonham (1963) and Elwood et al. (1974) both note the increased incidence of mortalities among babies presenting by the breech, they do not specify a causal relationship.

Pregnancy problems arise in **multiple births** because two or more infants are sharing an environment ideally designed for one. Commonly, one baby will develop faster than the other(s), commanding a greater share of the maternal resources. Although not usually on their own capable of reaching the normal weight for a full term singleton baby, the combined weight of twins or other multiple infants frequently exceeds this. Many authors, for example Morris and Heady (1955) and Peckham et al. (1982) have noted an association between multiple born infants and perinatal problems, multiplicity appears in fact to contribute to the exacerbation of many causes of death (Llewellyn-Jones, 1974).

Within the Leicestershire survey, records are kept of the type of case or control (whether singleton or multiple), the number of babies born in the current pregnancy and the number of stillbirths. Table 7.18 shows that many more cases than controls belonged to a multiple birth; of the 70 such cases, 67 were part of twins and 3 were one of triplets; all 10 multiple born controls were twins. The difference between the two groups is significant under the chi-square test.

Lambert (1976) has suggested that the likelihood of a stillbirth or infant mortality is increased if the baby is a boy. This view is tentatively supported with figures of 63 (61.2%) male and 40 (38.8%) female cases, compared with 52 (51%) male and 50 (49%) female controls; data for the remainder of births is missing. The chi-square statistic does not, however, indicate

significance for these few values. Peckham et al., (1982) have found that perinatal mortality rates are higher among like-sexed births than among twins of different sexes. This however cannot be tested here.

Season, day and time of birth have all at some time been hypothesised as having an association with perinatal or infant death (Adelstein et al., 1980). Month and hour of birth can be examined using the present survey but it is difficult to assess the validity of the results. On the one hand, by virtue of the selection process of the control population (being the next live delivery in the intended place of confinement of the case mothers) the time distribution of the controls is likely to reflect that of the cases; on the other hand, the particular subset of the data set being used here is not matched case/control set but rather a selection of the original data set based on the mother's place of residence. By deleting cases whose mothers live outside the Greater Leicester area this problem might be eliminated. Identification of a specific time period when babies seem to be more at risk is almost impossible. Between 11 a.m. and 3 p.m. there is an excess of controls over cases, but at other times the dominant group varies. Late afternoon, late evening and mid morning are periods when cases exceed controls but to ascribe any significance to this would be inappropriate. Similarly, the monthly variation in birth totals is apparently random.

In summation, birthweight, gestational age, presentation at birth and multiplicity have all been found to be significantly associated with perinatal death. These results support previous findings (Elwood et al., 1974; Butler et al., 1963; Peckham et al., 1982). In addition, a statistically significant association has been observed between type of labour and perinatal outcome.

Contrary to earlier evidence, the length of time spent in labour, the sex of the infant and the time of birth do not appear to be related here to perinatal mortality.

7.2.5 Health Care

Disease ecology and health care research have run parallel courses in medical geography. Arguably, health care may be considered as one of the factors within the mother's socio-medical environment which might have an effect on the outcome of her pregnancy. Certainly, past evidence (for example Butler and Bonham (1963), Elwood et al. (1974), and Kohl (1960)) points to a positive relationship between perinatal mortality and lack of adequate antenatal care. The perinatal survey addresses itself to the question of health care provision in several ways, in the main considering the frequency of different types of contact with the health services, such as in-patient hospital care, clinic attendance and visits to the G.P.

The first contact between mother and doctor is recorded twice in the survey; firstly, from the mother's medical records, and secondly, from her own estimate. These figures vary considerably, with mean values for cases and controls of 13.4 and 13.3 weeks from the medical records, and 9.3 weeks for both groups from the mothers' estimates. This discrepancy might be explained by the practice of using 'cooperation cards' to record all visits for antenatal care. These are given to mothers on their first hospital visit, normally at about 12 weeks gestation, and would not normally include records of visits to G.P.s prior to this. Frequency distributions for case and control groups are similar for both measurements. The conclusion must be that despite previous findings, (Kohl, 1960), the timing of a single, albeit the first, antenatal visit in Leicester is not crucial to the baby's well-being.

The pattern of subsequent antenatal care may however be more important. Butler et al., (1963) have suggested that in the absence of any antenatal care the risk of a perinatal mortality increases five-fold. The survey provides four measures of care.

The first measure is of the number of attendances for antenatal care, both at hospital and with the G.P., before 20 weeks gestation. Peckham et al. (1982) have suggested that early attendance for antenatal care decreases the mortality risk. The present study, however, cannot confirm this since the pattern of early attendances is very similar for both case and control mothers. The widest disparity occurs in the category of three attendances which comprises 25.9% of case mothers and 29.9% of control mothers. The difference between the two groups is not significant under the chi-square test.

A second count is of the number of attendances for antenatal care after 20 weeks. During the second half of pregnancy there is marked variation between attendance rates for case and control mothers. Control mothers make on average about ten visits for antenatal care over this period, compared with an average of six visits for cases. Table 7.19 shows how the frequency distributions for the two groups vary, the difference between them is statistically significant. Two factors may account for this. First, the number of attendances is related to the duration of the pregnancy. A mother who loses her baby at, say, 29 weeks gestation has less opportunity to seek care than the mother who carries her child to full term. Secondly, the difference may be due to a genuine inability or unwillingness among some case mothers to attend for antenatal care. It should be noted here that attendance for antenatal care is very much a voluntary activity and not under the control of a doctor. A key problem in matching health care

provision to those who most need it is that the people involved may be those who are less inclined or less able to seek it. Thus the lower attendance rate might reflect either a disinterest in, or inability to get to the care - both in themselves problems in the pregnancy.

The third measure of antenatal care is the number of days of in-patient care during the pregnancy. In contrast to out-patient care, mothers being admitted to hospital are usually there on the recommendation of a doctor. Thus, the most likely explanation of the higher average number of days care for case mothers (a figure of 6.99 days compared with 4.2 days for controls) lies in the doctors' practice of hospitalising women with pregnancy problems. Table 7.20 shows how many more case mothers received in-patient care over long periods. In-patient care can therefore be seen as a measure of the awareness, on the part of medical staff, of either existing or potential difficulties in the pregnancy. The causal relationship, if such exists, is in the opposite direction to that of normal, voluntary, attendance for care.

Finally, antenatal care is measured by the provision of a clinic session at the G.P.'s surgery or practice. According to Table 7.21 the mothers of controls have on the whole been better placed to receive antenatal care, assuming that they attended the G.P. and were made aware of the clinic session. It is reasonable to suppose that where care is easily accessible mothers will be more inclined to take advantage of it.

The type of antenatal care to be received by the mother is frequently decided at the start of the pregnancy. Initially the majority of women planned to receive shared consultant and G.P. care throughout pregnancy (see Table 7.22), some mothers planned full G.P./midwife care and a handful intended full consultant care. Just over 1% of all mothers planned to receive no care

Table 7.20

Number of days in-patient antenatal care	Cases		Controls	
	N	%	N	%
0- 2	387	84.1	458	67.0
3- 5	94	13.7	111	16.2
6-10	77	11.2	45	6.6
11-15	47	6.9	25	3.7
16-20	19	2.8	16	2.3
21-30	26	3.8	14	2.0
more than 30	35	5.1	15	2.2
	<u>685</u>	<u>100.0</u>	<u>684</u>	<u>100.0</u>

$\chi^2 = 34.3$ with 6 degrees of freedom
significance <0.001

Table 7.21

Provision of a clinic session at the G.P.	Cases		Controls	
	N	%	N	%
Yes	575	84.1	622	91.0
No	108	15.9	61	9.0
	<u>683</u>	<u>100.0</u>	<u>683</u>	<u>100.0</u>

$\chi^2 = 14.3$ with 1 degree of freedom
significance <0.001

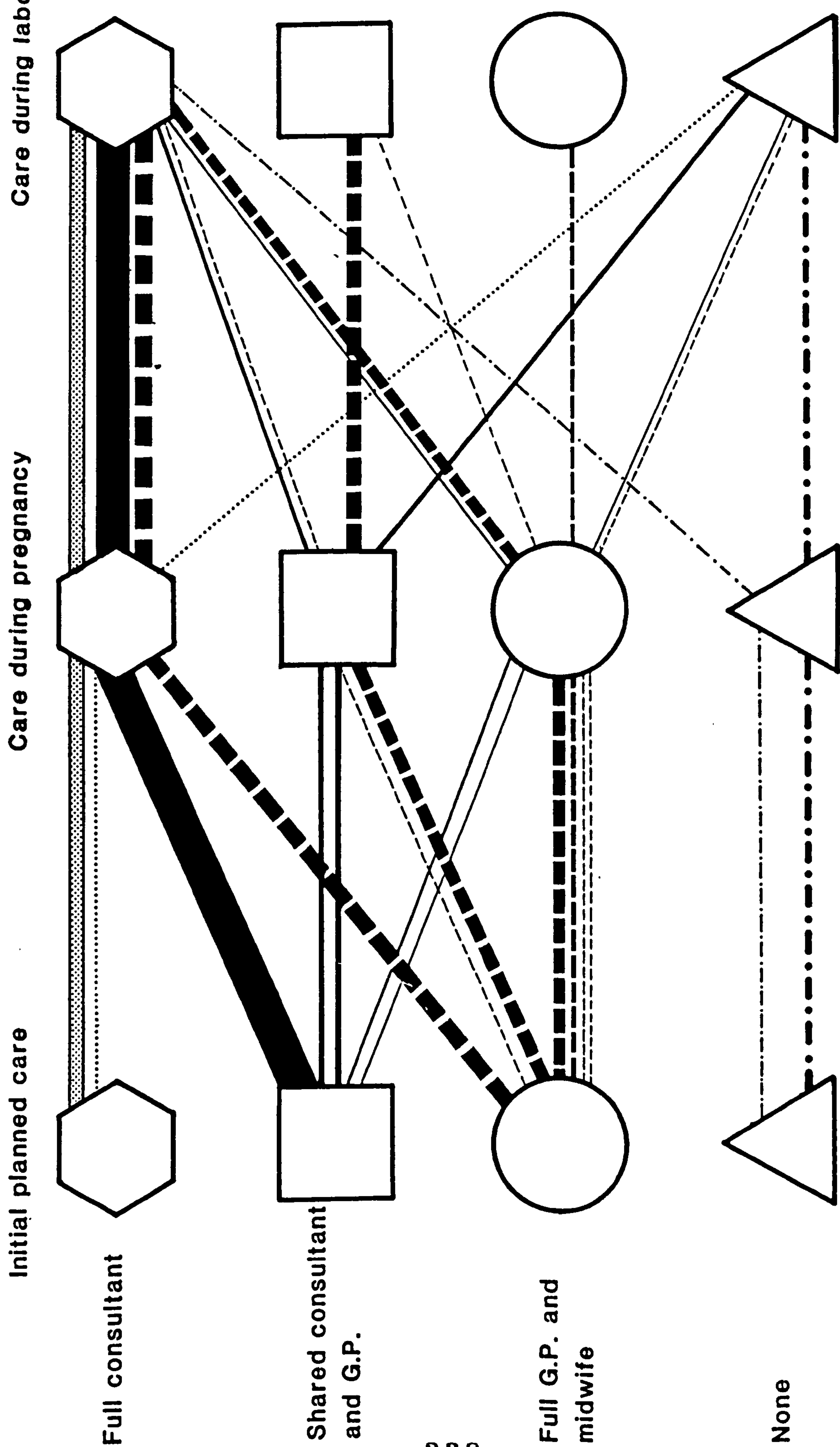
whatsoever. Although most mothers did not in fact change their plans as the pregnancy progressed, there is some evidence to suggest a movement to the more specialist type of care, especially among the mothers who were eventually to lose their babies (see Figure 7.2). Thus 86 (12.6%) case mothers changed to full consultant care during their pregnancy and a further 31 (4.5%) transferred to a consultant unit during labour. Corresponding figures for control mothers were 38 (5.6%) and 30 (4.4%) (see Figure 7.3). This is strongly suggestive of action being taken in response to problems in the pregnancy. A few changes to both combined G.P./consultant care and G.P./midwife care also occurred but there was little difference between cases and controls in this respect, and moreover, much change would not have been expected given the dominance of combined G.P./consultant care originally.

A possibly 'unplanned' change in care during labour is the unattended delivery. 11 (1.6%) case mothers fall into this category, indicating a strong element of risk to the infant. An unattended delivery is frequently associated with an unexpected arrival at home and the increased incidence of mortality among domiciliary confinements is well reported in the literature (see Macdonald Davies (1980)). Of the women in the survey, 19 (2.8%) lost their babies following a birth at home. Unfortunately it is impossible to standardise for this variable given the method of selection of control births - since only a handful of mothers intended to deliver at home, only five of the controls were taken from this group.

Butler and Bonham (1963) have suggested that the quality of antenatal care influences pregnancy outcome. Within the current survey, quality of care is measured by the level of the G.P.'s obstetric qualifications (see Clarke and Clayton (1983) for an

Figure 7.2

Change in care during pregnancy and labour : CASES

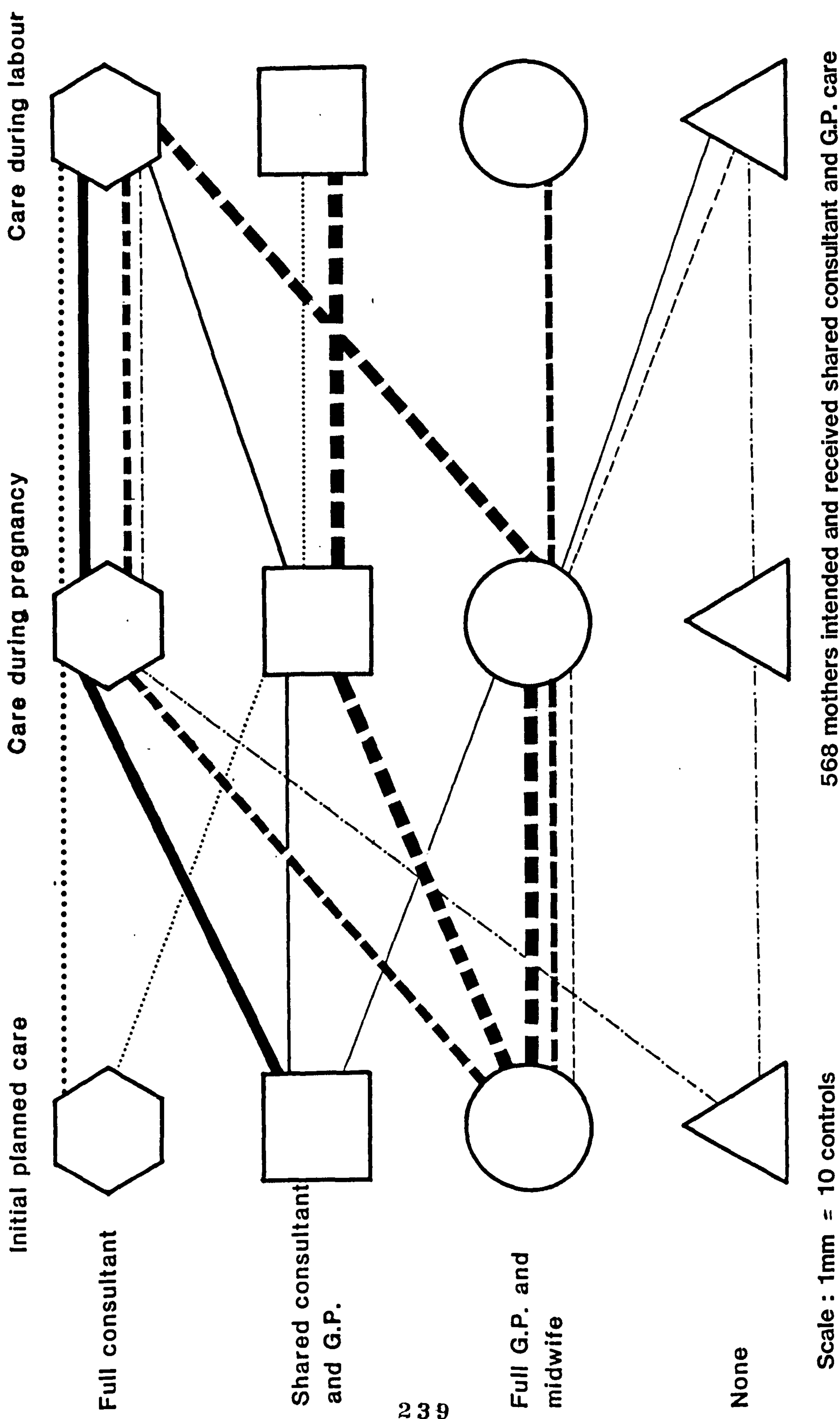


Scale : 1mm = 10cases

491 mothers intended and received shared consultant and G.P. care

Figure 7.3

Change in care during pregnancy and labour : CONTROLS



Scale : 1mm = 10 controls

568 mothers intended and received shared consultant and G.P. care

example of the use of this measure). Mothers being treated by G.P.s without formal obstetric qualifications are considered to be disadvantaged since it is possible that such G.P.s are less likely to identify abnormalities in the pregnancy immediately they occur. It is apparent from Table 7.23 that more case mothers were treated by G.P.s without obstetric qualifications than control mothers, moreover, the difference between the two groups is statistically significant.

The importance of health care during pregnancy and labour is widely recognised but relatively few authors have attempted to quantify the relationship. In this section the availability, quantity and quality of health care have all been shown to be significantly associated with perinatal outcome, particularly during the latter half of pregnancy. Interestingly, despite earlier evidence (Peckham et al., 1982), attendance for antenatal care early in the pregnancy does not appear from the present survey to be significant.

7.2.6 The character of perinatal death in Leicester

In addition to the many variables which might contribute to the infant's death, the Leicestershire perinatal mortality survey includes detailed information pertaining to the death itself. In particular, the stage of pregnancy at which death occurred, the primary cause of death and whether the death might have been avoidable are of interest.

Regarding the stage of pregnancy at which death occurred, just over half of the babies that died were stillborn (see Table 7.24) and most of these died during the antenatal period. Of the babies that were born alive, the greatest number died within the first day after birth, successively fewer died each day during the first week. It is apparent that the chances of a baby's survival are increased with its age.

Table 7.22

Initial planned care during pregnancy	Cases		Controls	
	N	%	N	%
Consultant/G.P.	564	82.5	594	86.8
G.P./Midwife	81	11.8	78	11.4
Consultant	27	3.9	9	1.4
No care	12	1.8	3	0.4
	—	—	—	—
	<u>684</u>	<u>100.0</u>	<u>684</u>	<u>100.0</u>

Table 7.23

G.P.'S obstetric qualifications	Cases		Controls	
	N	%	N	%
Not on the obstetric list	73	10.7	43	6.3
On obstetric list but no higher	435	63.6	455	66.5
On obstetric list with higher qualifications	176	25.7	186	27.2
	—	—	—	—
	<u>684</u>	<u>100.0</u>	<u>684</u>	<u>100.0</u>

$\chi^2 = 8.5$ with 2 degrees of freedom
significance <0.05

Table 7.24

Stage of pregnancy death occurred		Cases
~~~~~		
Stillbirths: Died during		
Antenatal period	267	
1st stage of labour	79	
2nd stage of labour	25	
	—	
	371	
Born alive, but died on		
Day 1	139	
Day 2	67	
Day 3	39	
Day 4	23	
Day 5	20	
Day 6	11	
Day 7	9	
	—	
	308	

Table 7.25

Was the death avoidable?	Cases	
	N	%
~~~~~		
Avoidable:		
1. Failure of antenatal care	50	7.3
2. Failure of care at delivery and/or labour	31	4.5
3. Poor patient compliance/failure to cooperate	24	3.5
4. 1 + 2	3	0.5
5. 1 + 3	6	0.9
6. 2 + 3	1	0.1
7. 1 + 2 + 3	0	0.0
9. Avoidable - neonatal	4	0.6
0. Refused amniocentesis if classified avoidable or refused termination	7	1.0
Unavoidable:		
8. Deaths where there has been a high quality of medical care and patient compliance	559	81.6
Deaths where the fetus was incompatible with life	—	—
	685	100.0

For each case in the survey the direct cause of death was recorded, as well as up to two antecedant causes. Those that occurred with high frequency are described below. The major direct causes of death to Leicester infants were fetal anoxia, fetal hypoxia, intra-uterine asphyxia and cerebral anoxia. Together these accounted for the death of 315 babies, or 46% of cases. These are all related conditions referring to a lack of oxygen to the infant at some stage during pregnancy or childbirth. They may be brought on by conditions in either the mother (e.g. severe pre-eclampsia, antepartum haemorrhage, heart disease or prolonged pregnancy) or through fetal causes, such as cord compression.

Prematurity, whatever the degree, was recorded as the direct cause of death for 108 infants (16% of cases). Prematurity normally results in the death of an infant when the baby is not developed or strong enough to adapt to extra-uterine life. Various causes of 'curtailed pregnancy' have been suggested, among these are included essential hypertension, malformation of the uterus, anaemia, acute infectious diseases (especially malaria), severe pre-eclampsia, accidental haemorrhage and rhesus incompatibility.

Anencephalus, with or without spina bifida, accounted for 40 cases (6%). Anencephaly is classified as a gross congenital malformation in which part of the skull is absent. It is one of the commonest central nervous system malformations and all infants so affected die within five days of birth. Among the possible causes are genetic and environmental factors; specifically, maternal disease (particularly viruses), inadequate maternal nutrition or, albeit rarely, maternal ingestion of a drug. Amniocentesis (the removal of some amniotic fluid for diagnostic purposes) is often performed to confirm this condition if a previous alpha feto-protein test has suggested it. The number of

babies reaching full term with anencephaly are only a proportion of the total who suffer, many are aborted early in pregnancy when the diagnosis is originally made.

Other congenital malformations account for a significant number of additional deaths: 'multiple malformations' was the direct cause of death for 28 infants (4%); malformations of the cardio-vascular system claimed 22 lives (3%); spina bifida, 12; hydrocephalus with spina bifida, 12; and chromosome abnormalities, 11. Contributory factors to these causes of death are listed under 'anencephalus'.

Respiratory distress syndrome was responsible for the death of 26 infants (4%). This occurs when there is a relative lack of 'surfactant' in the alveoli which impairs the lungs' capacity for gas exchange and hence oxygen intake.

Intra-uterine growth retardation, the direct cause of 23 (3%) deaths, is similar in effect to prematurity. Both result in low birthweight. The baby with retarded growth is normally referred to as a 'small-for-dates' baby. Growth may be retarded by a number of factors, although most relate to the ability of the mother to support the child. Multiple pregnancy, malformation and maternal disease are primary causes.

Antecedant causes of death were only recorded for about half the infants that died. The most common causes include premature labour (112 cases), premature separation of the placenta (50 cases), intra-uterine growth retardation (40 cases), pre-eclampsia/toxaemia (21 cases), premature rupture of the membranes (19 cases), induction of labour (17 cases) and prolapse or compression of the cord (10 cases). Some of these were also stated as tertiary causes of death, in addition multiple pregnancy (35 cases) and haemorrhage (19 cases) were also cited.

The last factor to be considered is whether the death was

avoidable or not. This is not a matter to be treated lightly and was only coded on the questionnaire following detailed discussions between all the medical personnel involved in the case. In Table 7.25 it can be seen that the majority of deaths were unavoidable. Of the 'avoidable' deaths, 38 (30%) in some way involved a lack of compliance on the part of the patient. 'Lack of compliance' usually meant refusal to attend an antenatal clinic at specified intervals, or a refusal of hospital admission when advised. (MacVicar et al., 1977).

7.2.7 Summary: Factors of significance in explaining perinatal mortality in Greater Leicester.

Perinatal mortality is undoubtedly a multi-factorial problem, but one which requires a prior understanding of each constituent element. This section has attempted to assess the significance of different survey variables to this problem in the light of previous findings. On the whole, previous work has been supported, with such factors as birthweight, gestational age, multiplicity and antenatal care all being found to be significantly associated with pregnancy outcome. On the other hand, some variables previously suggested as important associative agents have not been found to be significant within this data set. These include marital status, early contact with medical services, maternal age and smoking habits.

Table 7.26 summarises the main findings of this section. All variables for which a case-control comparison yielded a significant chi-square value have been included, these are deemed to be key indicators for the remainder of this work. It should be noted however, that any conclusions based on a statistical test of significance are probabilistic (Silk, 1979). Thus, although the hypothesis of independence is rejected at the given levels of

Table 7.26

Variables associated with perinatal mortality : first summary of results

Variable	χ^2	Degrees of freedom	Significance of χ^2 value (ie. proba- bility that H_0 is correct)	N (number of observa- tions)

1. Mother's physical characteristics				
Maternal height	10.47	5	<0.1	1360
2. Mother's obstetric history				
Parity	16.41	6	<0.05	1369
Gravidity	12.75	6	<0.05	1368
Previous pregnancy outcome:				
Alive/dead	7.17	1	<0.01	856
Alive and well/other	47.7	1	<0.001	856
3. Socio-economic characteristics				
Mother's social class	13.57	4	<0.01	1366
Mother's socio-economic group	13.57	4	<0.01	771
Father's socio-economic group	11.48	4	<0.05	1226
Mother's ethnic group	15.5	5	<0.01	1369
Mother's region of birth	15.2	3	<0.01	1369
Mother's religion	9.5	3	<0.05	472
3. Infant characteristics				
Birthweight	655.1	6	<0.001	1369
Gestational age	512.5	2	<0.001	1369
Type of labour	155.5	7	<0.001	1368
Presentation at birth	148.1	2	<0.001	1360
Multiplicity	47.7	1	<0.001	1369
5. Health care				
Attendance for ANC after 20 weeks gestation	364.6	4	<0.001	1351
In-patient ANC	34.3	6	<0.001	1369
Provision of clinic at G.P.	14.3	1	<0.001	1366
G.P.'s obstetric qualification	8.5	2	<0.05	1368

significance, we cannot be certain that this is necessarily reflecting a true process. For this reason, all variables shown to be significant at the 0.1 level or above, have been included although in fact a significance level of 0.05 is more appropriate.

The relationships between these variables and adverse perinatal outcome will be explored further in the next chapter.

7.3 Leicestershire Births Records : description of data set

In 1980 there were 6,302 births to mothers residing in the Greater Leicester study area. Of these, 6,192 lived, 46 were stillborn and 57 died within four weeks of birth. According to the perinatal survey, there were 95 perinatal deaths in that year. The excess of 8 mortalities in the total births data is probably accounted for by deaths occurring between one week and four weeks after birth, that is, in the late neonatal period.

Six additional variables record **birth outcome** dichotomously, these are displayed in Table 7.27. It can be seen that 38 babies were stillborn in 1980 and a total of 34 more died within one week. These numbers fall short of those in both the perinatal survey and elsewhere in the births data, most probably this is because of the number of births that are counted as 'missing'. Since any analysis of birth characteristics should include all births, whatever their outcome, and already we have a comprehensive and reliable data set for the mortalities in which we are interested, these discrepancies must be seen in the context of the full births data. As such they constitute a very small and insignificant proportion.

The average age of all mothers in this data set was approximately 25.5 years, slightly younger than that indicated by the perinatal survey. The distribution across the age groups is shown in Table 7.28. Compared with the mothers interviewed in the perinatal survey, this table shows smaller proportions in both the

Table 7.27

Birth outcome	Total births		
	'Yes'	'No'	Missing
Mother transferred to hospital	1	530	5771
Alive at birth	493	38	5771
Stillbirth	38	493	5771
Died within 24 hours of birth	21	510	5771
Died within 6 days of birth	13	518	5771
Died within 27 days of birth	5	526	5771

Table 7.28

Maternal age	Total births	
	N	%
16 or less	60	1.0
17 to 19	644	10.2
20 to 24	2184	34.7
25 to 29	2070	32.8
30 to 34	1024	16.2
35 or more	319	5.1
	<hr/>	<hr/>
	<u>6301</u>	<u>100.0</u>

Table 7.29

Ethnic group	Total births	
	N	%
African	34	0.5
British Isles	4601	73.0
European (excluding British)	51	0.8
Indian - from India	689	11.0
Indian - from Uganda	138	2.2
Indian - from Kenya	304	4.8
Indian - from Africa (excluding Uganda and Kenya)	191	3.0
Pakistani	59	0.9
West Indian	61	1.0
Others	174	2.8
	<hr/>	<hr/>
	<u>6302</u>	<u>100.0</u>

lowest and highest age groups - among the case mothers 1.6% were aged 16 or less and 7.7% were aged 35 or more (see Table 7.1). Interestingly, both case and control groups had the largest proportion of mothers in the 25 to 29 year category as opposed to the 20 to 24 year class shown in Table 7.28. If cases are compared with total births using the chi-square test the resulting value of 16.61 with 5 degrees of freedom is much larger than that obtained by the simple case/control comparison. It is in fact significant at the 0.001 level. This association has been previously identified by several authors (e.g. Butler and Bonham, 1963; Elwood et al., 1974) but surprisingly was not detected by the perinatal mortality survey.

A slightly different classification system was used for coding the births data into ethnic groups compared with the survey data. Ten different groups were distinguished, as shown in Table 7.29. The distribution shown here more closely resembles the proportions of control mothers in each ethnic group than case mothers. Most notably, although 73% of all mothers delivering in 1980 were in the "British Isles" group, only 68.9% of case mothers could be classified as such (UK and Eire categories combined). Conversely, a total of 27.9% of case mothers belonged to the Asian ethnic groups, compared to only 21.8% of all mothers; and 1.9% compared with 1.0% were West Indians. All of these observations were made using control population alone.

Mean values for birthweight and gestional age are 3199.5g and 39.2 weeks respectively, these both being similar but slightly lower than the values for the control population. Frequencies which may be compared with those in the previous section are shown in Tables 7.30 and 7.31.

The larger data set contains a higher proportion of babies

Table 7.30

Birthweight (grammes)	Total births	
	N	%
500 or less	5	0.1
501 to 750	9	0.1
751 to 1000	25	0.4
1001 to 1250	18	0.3
1251 to 1500	22	0.3
1501 to 1750	45	0.7
1751 to 2000	75	1.2
2001 to 2250	140	2.2
2251 to 2500	275	4.4
2501 to 3000	1484	23.5
3001 to 3500	2342	37.2
3501 to 4000	1455	23.1
4001 and over	407	6.5
	<u>6302</u>	<u>100.0</u>

Table 7.31

Gestational age	Total births	
	N	%
Pre-term (20 to 36 weeks)	468	7.4
Term (37 to 41 weeks)	5505	87.4
Post term	327	5.2
	<u>6300</u>	<u>100.0</u>

of low gestational age (ie. pre-term infants) and of low birthweight than the survey's control population. This is to be expected, given that the births data include stillbirths and babies that were to die in the neonatal period. The non-survivors, however, are not numerous enough to markedly alter the overall distribution of either variables in the way that the case group alone has.

Now that it is possible to examine a large number of births, the proportion of survey control babies born post term seems surprisingly high at 9.7% compared with 5.2% of all births.

Applying the chi-square test to cases and total births data for both of these variables gives exceptionally high values for the chi-square statistic. Namely, 2776.7 and 12 degrees of freedom for birthweight; and 1682.9 with 2 degrees of freedom for gestational age. These are clearly both far in excess of the 0.001 significance level.

The largest multiple birth to a Leicester mother in 1980 comprised three infants. One set of triplets, 55 pairs of twins and 6,189 singleton births constitute the full data set. Proportionately, the control group accurately reflect the overall population, with 1.5% multiple births compared to 1.7% overall. Both of these figures are notably less than the equivalent for the survey case group, where 10.2% of babies were part of a multiple birth.

Regarding infant sex, of the 6,302 births in total, 3,234 (51.3%) were male, 3,065 (48.6%) were female and 3 (less than 0.1%) were indeterminate. These proportions are similar to those of the control population. From cross-tabulating infant sex and pregnancy outcome (live/stillborn/dead), there is evidence that survival among male babies is less than among females, with 1.8% of boys dying within four weeks and only 1.4% of girls. This

Table 7.32

Total births and perinatal deaths : summary of results

Variable	χ^2	Degrees of freedom	Significance of χ^2 value (ie. proba- bility that H_0 is correct)	Number of cases	Number of births
Maternal age	16.6	5	<0.001	685	6301
Ethnic group		Not directly comparable			
Birthweight	2776.7	12	<0.001	685	6302
Gestational age	1682.9	2	<0.001	685	6300
Multiplicity	172.59	1	<0.001	685	6302
Infant sex	3.95	1	<0.05	103	6299

difference is very much less than that suggested by the short run of figures from the perinatal survey. The difference between the proportions of each sex among cases and total births is still significant under the chi-square test.

In summary, these variables are all that can be extracted from the total births data set to be used for comparative purposes, so it is impossible therefore to fully validate the control population. Nevertheless, of the six variables which could be compared, the control population correctly identified the direction of the relationships in five, these being ethnicity, birthweight, gestational age, multiplicity and infant sex. It was unsuccessful in establishing a significant association between maternal age and pregnancy outcome.

7.4 The distribution of perinatal mortality in Greater Leicester

7.4.1 Methods for combining data from different sources

The three data sets used thus far to describe Greater Leicester - census, perinatal survey and births records - have two factors in common: the global extent of their coverage and the small size of the areal units to which they refer. They differ, however, in the delimitation of these areal units. As noted in chapters 4 and 6, the census data are made available on the basis of enumeration districts and have subsequently been aggregated into 'new small areas', while perinatal and births data are referenced to individual postcodes. The problem encountered here is to reconcile these two spatial referencing systems.

The location of each NSA is held on computer in two formats: as a string of coordinates delimiting the areal boundary, and as a mathematical centroid derived from this string. The postcode, on the other hand, is represented by only a single point, located at the start of the postman's walk. This results in a problem of

generalisation, since the point inevitably refers to an extreme location within a postcode area which is itself of an irregular shape. As the only definition available, however, the use of the point representation for postcodes is justified; for the NSAs such generalisation cannot be tolerated, therefore for these units area boundaries must be used.

There are three methods available for matching geographically point (postcode) and area (NSA) data. First, the data sets may be mapped independently for subsequent visual comparison. Being simple in execution, this method is intuitively attractive albeit subjective. It is particularly suited to a preliminary descriptive analysis and will be used in this context within the present study (see section 7.4.3). Secondly, the two spatial units may be assigned to a third, quite different, areal unit - for example, the grid square. As long as the square is large enough to avoid excessive overlap across cell boundaries of either postcode or NSA, this method is applicable. In practice, the choice of resolution of the grid squares and the acceptability of the use of an arbitrary spatial unit are obstacles. The third method entails allocating individually the smaller of the spatial units to the larger, in this case the postcodes to the NSAs. This method is described below.

Manually, the process of determining in which NSA each postcode lies is straightforward, although time-consuming and prone to human error. It involves plotting postcode coordinates on a base map showing new small area boundaries. To perform this task automatically requires that both postcodes and area boundaries should be held in machine-readable form, and also that a suitable computer algorithm to perform the 'point-in-polygon' test should be available. Program MPPTPL (Appendix 2.2) was used for this task. Variations of this program were also used to create an index

of postcodes to EDs; census wards, census districts and parishes. In theory the algorithm could cope with any combination of points and digitised area boundaries, limits are set only by computing time and memory constraints.

7.4.2. Allocation of births and mortalities to clusters and census areas - statistical comparison.

When the original EDs were aggregated into NSAs (see Chapter 5) two of the stated purposes of the classification were first to generate, and second, to enable the testing of hypotheses regarding the different characteristics of the constituent areas. This section uses the classification to explore the relationship between the geographical distributions of births and perinatal mortalities within the study area. The general hypothesis is that perinatal deaths are not evenly distributed with respect to total births, and that those clusters which might be described as 'deprived' in a socio-economic sense will also suffer the worst perinatal experience.

Using the 'point-in-polygon' approach described above, perinatal mortalities and births were allocated to EDs and NSAs. Both of these spatial units belong to hierarchies which may be identified from the smallest area code. Thus the ED named '32MMAA01' is ED number '01' which lies within ward 'AA', district 'MM' and county '32'. Similarly, new small area '1AA' is the area labelled 'AA' within cluster '1'. By assigning observations to the smallest areas in each hierarchy they are by default allocated to the larger zones.

Table 7.33 shows the proportion of total births, survey cases and survey controls in each cluster, the number of EDs in each cluster is given as an approximate indication of cluster size. Inspection of this table suggests that there are evident

Table 7.33

Analysis by cluster									
Cluster Code	Cluster Label	Enumeration Districts		Total Births		Cases		Controls	
		N	%	N	%	N	%	N	%
1	Established private residential	97	11.5	675	10.7	46	6.7	68	9.9
2	Low status, council	141	16.6	1118	17.7	124	18.1	108	15.8
3	Upwardly transient	35	4.1	209	3.3	36	5.3	26	3.8
4	Low status, ethnic quarter	107	12.6	1134	18.0	179	26.1	128	18.7
5	High social status, low density	167	19.7	1323	21.0	123	18.0	160	23.4
6	Socially dependant, council	45	5.3	290	4.6	28	4.1	29	4.2
7	High socio-economic	132	15.6	723	11.5	55	8.0	85	12.5
8	Elderly population	53	6.3	308	4.9	41	6.0	34	5.0
9	Private renting	70	8.3	522	8.3	53	7.7	36	6.7
		847	100.0	6302	100.0	685	100.0	684	100.0

Total births and cases:

$\chi^2 = 49.4$ with 8 degrees of freedom, significance <0.001

Cases and controls:

$\chi^2 = 27.87$ with 8 degrees of freedom, significance <0.001

Table 7.34

Analysis by Census District						
District	Total births		Cases		Controls	
	N	%	N	%	N	%
Blaby	617	9.8	49	7.2	86	12.6
Charnwood	513	8.1	38	5.5	55	8.0
Harborough/Hinckley + Bosworth	78	1.3	9	1.3	7	1.0
Leicester City	4537	72.0	537	78.4	463	67.7
Oadby + Wigston	557	8.8	52	7.6	73	10.7
	6302	100.0	685	100.0	684	100.0

Total births and cases:

$\chi^2 = 14.39$ with 4 degrees of freedom, significance <0.01

Case and controls:

$\chi^2 = 22.5$ with 4 degrees of freedom, significance <0.001

differences between the distributions of cases and total births, most notably in the cluster labelled 'low status ethnic quarter'. This cluster comprises 26.1% of cases, but only 18% of all births. When the chi-square statistic is calculated for these data a significant value of 49.4 is obtained. Statistically speaking, this result should be treated with care since almost half of this value results from the difference between birth and case values in the aforementioned 'ethnic' cluster. Of lesser importance are the relative excesses of births over cases in the 'indistinct' and 'high socio-economic status' clusters, and of cases over births in that labelled 'transients'.

If, instead of comparing cases to all births, they are compared with controls, these differences are not as marked. Proportionately, the controls are closer to the overall population than to the cases; this was to be hoped for, but not necessarily expected given the nature of the control selection process. The largest differences between the control and overall populations lie in clusters 2 (low status, council), 5 (high social status, low density) and 9 (private rented), none of which contributed significantly to the chi-square value above. The chi-square value for the case-control comparison is a significant 27.87.

In Table 7.34 the births, cases and controls are allocated to census districts (shown in Figure 7.4). The two smallest districts in this survey have been combined. As before, the distribution of controls more closely resembles that of total births than that of cases although in most districts the former occupies a median position.

Examination of individual districts shows Leicester city to have a substantially higher proportion of cases than either controls or total births, and Blaby and Charnwood districts to have less. The combined Harborough and Hinckley and Bosworth

districts have few births and contribute little to the calculated chi-square values.

Lambert (1976) and Spicer and Lipworth (1966) have both previously found the incidence of perinatal mortality to be related to geographical location as measured on a rural/urban dichotomy. Thompson (1968) suggested a multivariate solution which included both level of industrialisation and density of population among the associative variables. All of these findings can be tentatively supported by evidence from Leicester, since the city is undoubtedly the most 'urban', 'industrialised' and densely populated part of the study area; the other districts mainly constitute the suburban fringe.

In conclusion, the statistical comparison of the distributions of births and perinatal mortalities has gone some way to confirming the hypothesis of multiple deprivation. Parts of the study area with proportionately more perinatal deaths include the new small areas belonging to the 'ethnic' cluster, itself characterised by large households, overcrowding, high rates of migrancy and lack of basic amenities (see Chapter 5); and the most densely populated city district. These findings need not be independent.

Although an attempt was made to analyse the distribution of perinatal mortality by census ward, the results have not been reported here. A total of 50 wards and part wards are included in the study area (see Figure 7.4), too many to justify the use of the chi-square test.

Of the two aggregations, the analysis by cluster produced the more significant chi-square results, probably because of the greater internal homogeneity of these areas. This clearly demonstrates the validity of the earlier classification.

7.4.3 Cartographic comparison of births and mortalities

The process of allocating births and survey data to geographic units enables an analysis of perinatal mortality at an aggregate level. In Figures 7.5 and 7.6 these observations are plotted individually. Figure 7.5 shows the distribution of perinatal deaths across Greater Leicester; digits represent the number of deaths over the seven year period, 1976 to 1982, in each postcode area. Postcodes having no mortalities are not plotted.

It can be seen from this map that the distribution of mortalities across the study area is far from uniform. The greatest concentrations are to be found close to the city centre especially in the Highfields and Belgrave areas, and to a lesser extent towards West End. The highest values for individual postcodes are found at Syston, having a total of nine deaths, and Kirby Muxloe, with five. These are explained in part by the larger size of their respective postcode areas, it is noted that in both examples there are no other mortalities in the immediate vicinity. In addition, there is a possibility that higher perinatal death rates might be associated with the older village centres. This will be examined later.

Figure 7.6 shows the distribution of all mothers giving birth in 1980. The method of display is not entirely satisfactory, but it is constrained by the large number of births and the small size, and proximity, of the postcode areas to which they are referenced. The greatest number of births are seen in a north-south ellipse around the city centre covering such areas as Belgrave, West End, Aylestone, Highfields and Spinney Hills.

In order to standardise the perinatal mortalities using the births data, a perinatal mortality rate has been plotted. (Figure 7.7) Based once again on aggregations of these data into NSAs, the perinatal mortality rate is calculated as the ratio of all

Figure 7.5

Perinatal mortality – cases

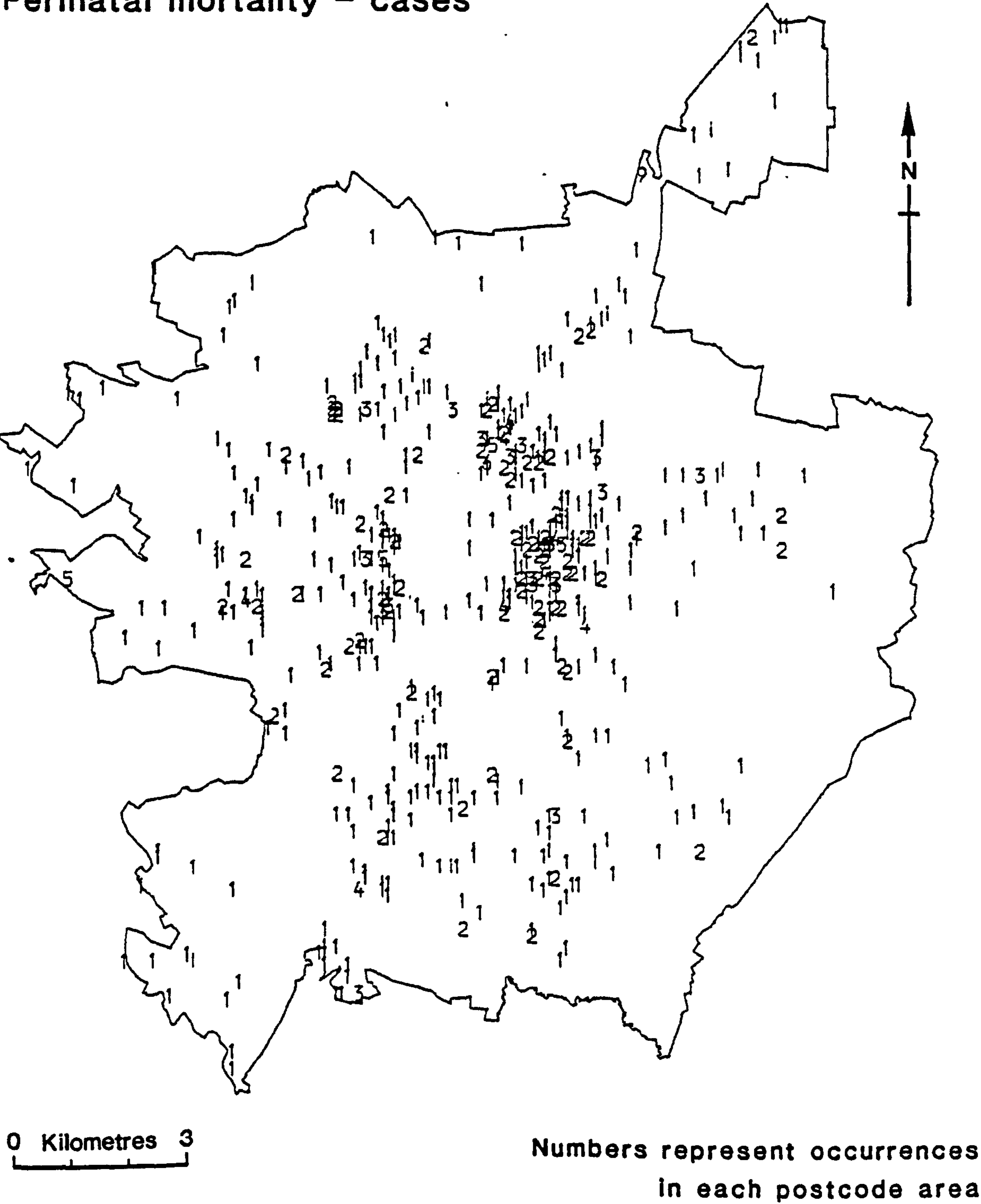


Figure 7.6

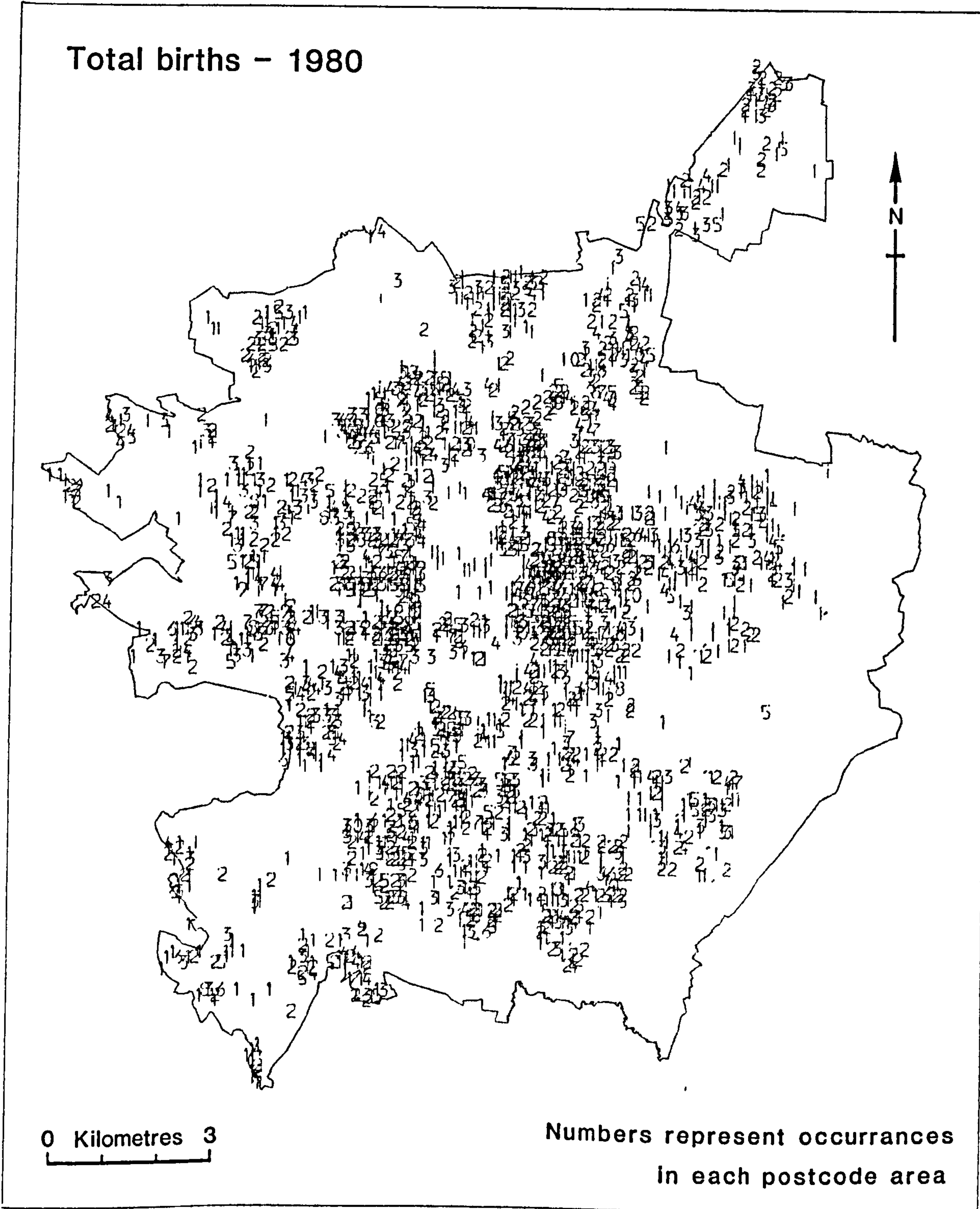
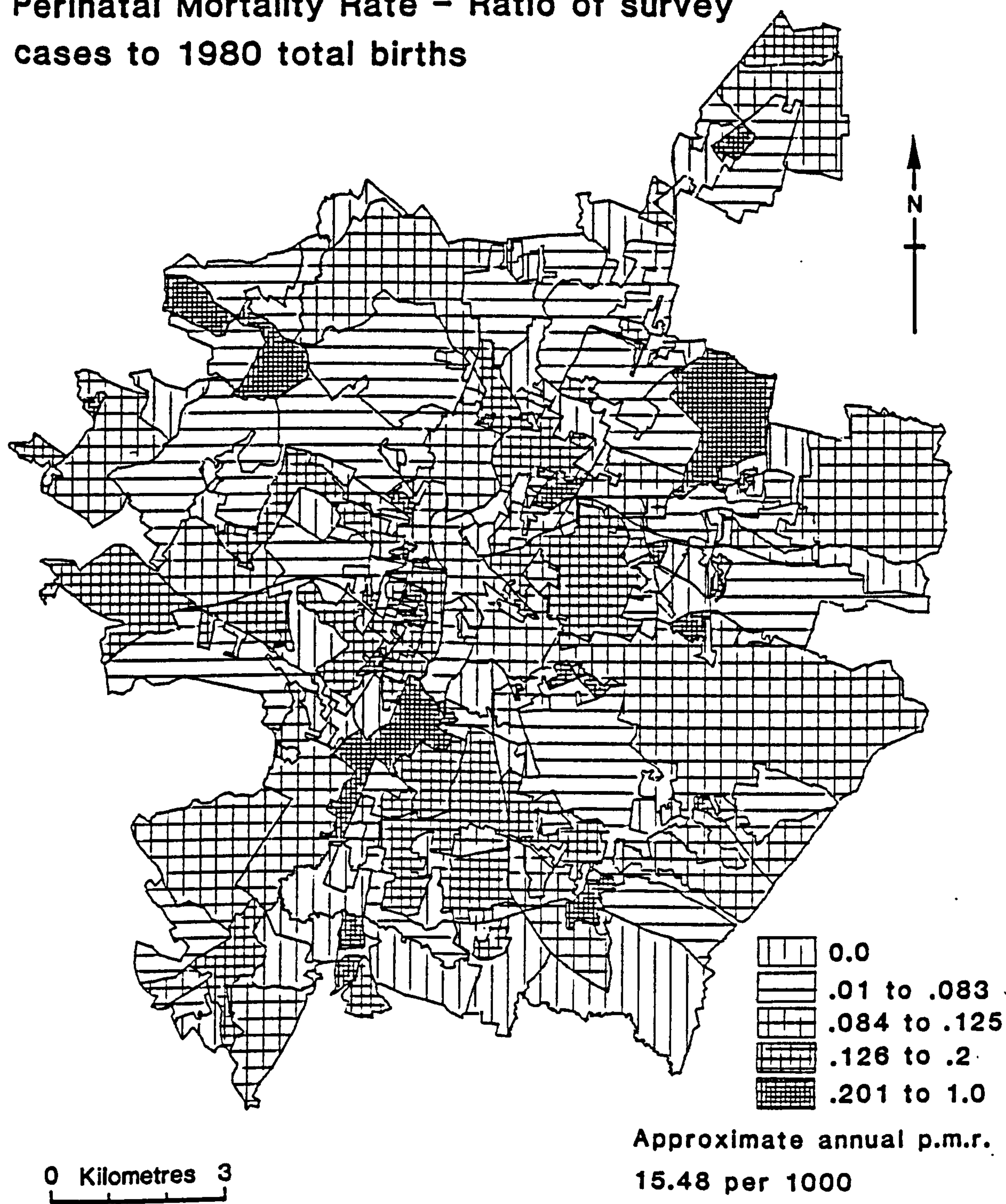


Figure 7.7

Perinatal Mortality Rate – Ratio of survey cases to 1980 total births



PLOTTED BY >> CHORO << 84/05/09.
LEICESTER UNIVERSITY

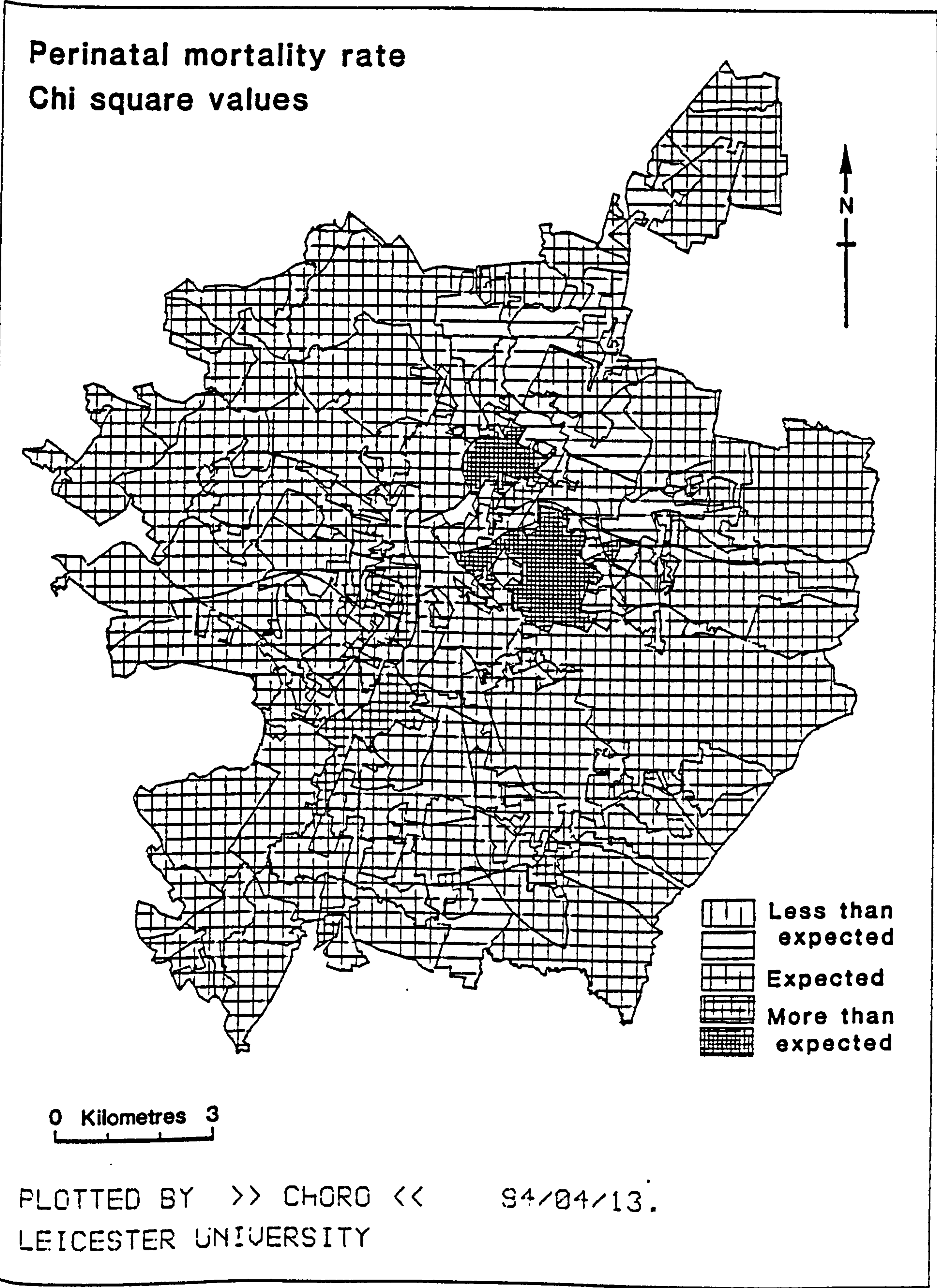
survey cases over the seven year period to total births in 1980. This of course gives a figure greatly in excess of the conventional annual perinatal mortality rate, therefore an approximation of the latter is given as a guide at the foot of the map.

The pattern shown in Figure 7.7 highlights certain of the smaller, inner-city NSAs, no doubt the result of mapping ratio values for small numbers, as well as isolated suburban areas. Viewed in conjunction with Figures 7.5 and 7.6, parts of West End, Aylestone, Braunstone, New Parks, Belgrave, Spinney Hills, Highfields, Kirby Muxloe, Anstey and Syston all fall into the top two categories and could be deemed to have the highest real rates.

Following the example set by Chapter 6 in plotting chi-square values for each variable, these values were calculated and plotted for the perinatal mortality rate. In this case, the overall mean value against which each area was compared was simply the ratio of mortalities to births, while expected values were calculated in relation to births and not to total population as before. Thus the resultant map (Figure 7.8) shows areas with both good and severe mortality experience after having standardised for the pattern of births. Two zones stand out as having abnormally high perinatal mortality rates - Highfields and Belgrave. These areas have already been identified as having the characteristics of multiple deprivation, this further finding confirms this. Several other areas also have a higher rate than expected, these have been mentioned before and include West End, Clarendon Park and Wigston town centre.

A few areas have lower than expected perinatal mortality rates, namely Birstall, Thurmaston, West Humberstone and Southeast Wigston. The remainder of the study area falls into the

Figure 7.8



central 'expected' category.

Finally, a caution for the interpretation of this map: the calculation of the chi-square statistic for mapping ideally demands a minimum number of observations as the expected value for each area. This constraint is not met here and the effect of this on the map is difficult to determine, hence the need to examine all the maps of perinatal mortality together.

Summarising, the mapped distributions do reflect the findings of the statistical analysis, at least on an aggregate level. Both absolutely and relatively the ethnic areas of Highfields and Belgrave are the most disadvantaged in terms of their mortality experience. Other socially handicapped areas such as the council estates at Aylestone, Braunstone and New Parks; the older village centres of Kirby Muxloe and Syston; and the congested inner city area of West End have above average levels of perinatal death.

7.5 The distribution of risk

In section 7.2 an association was found between perinatal outcome and some twenty variables from the perinatal survey. Based on both the tables included in this section and the previous findings of other authors, Table 7.35 summarises those categories of each variable considered to be most highly associated with perinatal death. These are defined as "risk categories". The remainder of this chapter is concerned with describing and explaining the distribution of survey members exhibiting these characteristics. Both case and control mothers are included since in this context both groups are considered to be equally "at risk".

For current purposes the distributions plotted may be divided into three groups; those that are dominated by the Highfields and Belgrave clusters, those that follow the overall

Table 7.35 Summary of risk categories

Variable	Categories of Greatest Risk	Table Number (for reference)

Maternal height	150cm. or less	7.2
Parity	0, 4 or more	7.3
Gravidity	1, 7 or more	7.4
Previous pregnancy outcome	Dead, Not alive and well	7.5, 7.6
Mother's social class	4, 5, housewife	7.7
Mother's s.e.g.	1, 11	7.8
Father's s.e.g.	7, 10, 11, 15	7.8
Mother's ethnic group	Indian, West Indian	7.11
Mother's region of birth	Africa, Asia	7.12
Mother's religion	Hindu	7.13
Birthweight	2000g or less	7.14
Gestational age	36 weeks or less	7.15
Type of labour	Medical induction	7.16
Presentation at birth	Breech	7.17
Multiplicity	2 or more infants	7.18
Attendance for ANC after 20 weeks gestation	5 visits or less	7.19
In-patient antenatal care	20 days or more	7.20
Provision of clinic at G.P.	No antenatal clinic	7.21
G.P.'s obstetric qualification	Not on obstetric list	7.23

pattern of severe mortality experience as identified in the previous section, and the remainder which do not fall into either of these classes.

Seven maps are dominated by clusters in Highfields and Belgrave. Half of these are categories of ethnic or religious variables, namely, Indian (Figure 7.18), African and Asian (Figure 7.19), West Indian (Figure 7.20) and Hindu (Figure 7.21) mothers. For these, the clustering is expected, following closely the distribution of persons born in the New Commonwealth or Pakistan (see Figure 6.10). The fifth map, showing mothers of height less than or equal to 150 cm. (Figure 7.9) also reflects the distribution of the immigrant population. The remaining two variables are concerned with the provision of primary health care - the availability of an antenatal clinic session at the G.P.'s surgery (Figure 7.32), and the G.P.'s level of obstetric qualification (Figure 7.33). That these two variables exhibit similar patterns is to be expected since a G.P. without specialised knowledge of obstetrics is less likely to provide an antenatal clinic than one with a specialist qualification. It is also predictable that the mothers attending these doctors will be spatially clustered since patients normally register with a G.P. convenient to their home. It is particularly unfortunate, therefore, that the doctors choosing to practice in the areas with the greatest numbers of births should be relatively less qualified to deal with these.

Eight maps fall into the second category, having distributions similar to the map of all cases (see Figure 7.5). Thus, Highfields and Belgrave are still important clusters, but in addition, the areas of West End, New Parks, Braunstone and Aylestone/Saffron Lane have above average incidences. Figure

7.25, showing the distribution of pre-term infants, is predictably similar to the overall distribution of cases since prematurity is a major cause of adverse perinatal outcome. The distribution of breech presentations (Figure 7.27) has a less direct relationship with perinatal mortality but is probably explained by the pattern of pre-term births (see Section 7.2.4). For both of these variables cases greatly outnumber controls, a characteristic which does not apply to the maps of socio-economic variables. The distributions of survey parents in the partly skilled and unskilled occupational groups (Figures 7.14, 7.15 and 7.17) are similar to the patterns identified by the census (see Figures 6.25 and 6.26). As previously discussed, social 'deprivation' is shown to be accompanied by higher perinatal risk. The sixth map in this group shows mothers making between 1 and 5 visits for antenatal care after 20 weeks gestation (Figure 7.30), a relatively low figure compared to the norm. The likelihood that a mother attends for antenatal care depends both on her level of accessibility (determined by such factors as physical distance to the nearest clinic and the availability of transport and of 'babysitting' facilities for older children if necessary) and on her expectation of that care. Both accessibility and expectation may be expected to diminish with socio-economic status. Lastly, the two maps showing previous perinatal outcome (Figures 7.12 and 7.13) exhibit distributions similar to the current pattern of mortalities. This pattern might have been predicted, given that the perinatal problem has been shown to have a spatial component and that the underlying factors contributing to this have not substantially changed within recent years.

The final set of maps have distributions not immediately similar to the overall pattern of cases. Two of these are large 'general' categories of variables; namely, housewives and zero

parity mothers. **Housewives** are predictably shown in some numbers throughout the study area (Figure 7.16). Assuming the perinatal survey supports the census findings, the expected distribution of these mothers would be the inverse of Figure 6.32 which shows married women at work. This appears to be the case. However, the limitations associated with using the mother's social or socio-economic class as an explanatory variable were noted in Section 7.2.3, and it is suggested that such limitations also apply to any interpretation of the mapped pattern. **Zero parity** mothers are also uniformly distributed across the study area (Figure 7.10), a pattern which might be anticipated since it is self-evident that all mothers will at some stage fall into this category. Although a behavioural hypothesis might be offered regarding life cycle changes and the distribution of new mothers, the classic movement to the suburbs assumes both that mothers have a free choice in where they live (particularly unlikely in the rented sector) and that they exercise this choice prior to the birth of the first child. Neither of these assumptions can be supported here.

The remaining maps have comparatively few observations plotted and are therefore difficult to liken to the previous distributions. For example, although some clustering is evident around Highfields for both **high parity mothers** (Figure 7.11) and **multiple births** (Figure 7.28), and moreover minor clusters can be identified in North Braunstone (high parity) and West End (multiple births), elsewhere the few observations are scattered and difficult to substantiate. This applies also to the distributions of **medically induced labour** (Figure 7.26), **mothers receiving no antenatal care after 20 weeks gestation** (Figure 7.29) and **mothers receiving over 20 days in-patient antenatal care** (Figure 7.31). It must be concluded therefore, that these

variables have no spatial component.

Three maps have not been included in the above discussion - those showing the distribution of low birthweight infants. The lowest of the three categories, birthweight less than 1000g, includes infants which almost certainly were delivered prematurely and whose life expectation must have been low. Clusters of these infants can be seen around West End, Belgrave, Highfields and Aylestone (Figure 7.22), although it is unlikely that this distribution is statistically significant. The 1001-1500g birthweight category is again dominated by cases but in contrast, little clustering is evident (Figure 7.23). The final low birthweight group, 1501-2000g, is interestingly the smallest of the three. Although cases, and some controls, are scattered throughout the study area greater concentration is shown than in the previous two maps (Figure 7.24). Two factors are of interest in interpreting this pattern. Firstly, babies in this birthweight group are no longer small enough for birthweight alone to necessarily determine infant outcome; and secondly, this group is likely to contain a higher proportion of 'small for dates' babies than the lower birthweight categories. In explaining the mapped pattern therefore, non-medical factors such as socio-economic disadvantage and ethnicity will assume a greater importance.

Two conclusions can be drawn from the analysis in this section. Firstly, where variables are common to both the perinatal survey and the census there is a close correspondence between the maps from the two sources. This has implications for both extending the current work by examining additional census variables, and for applying similar techniques to other areas. Secondly, the mapping exercise has emphasized the interdependence of many of the variables - see for example the similarity between the maps of the socio-economic variables and those showing

attendance for antenatal care. There is of course no statistical 'proof' of a causal relationship but the maps continue to lend support to previous findings.

Figure 7.9

Perinatal Mortality – Risk Characteristics

Mothers of height less than
or equal to 150 cm

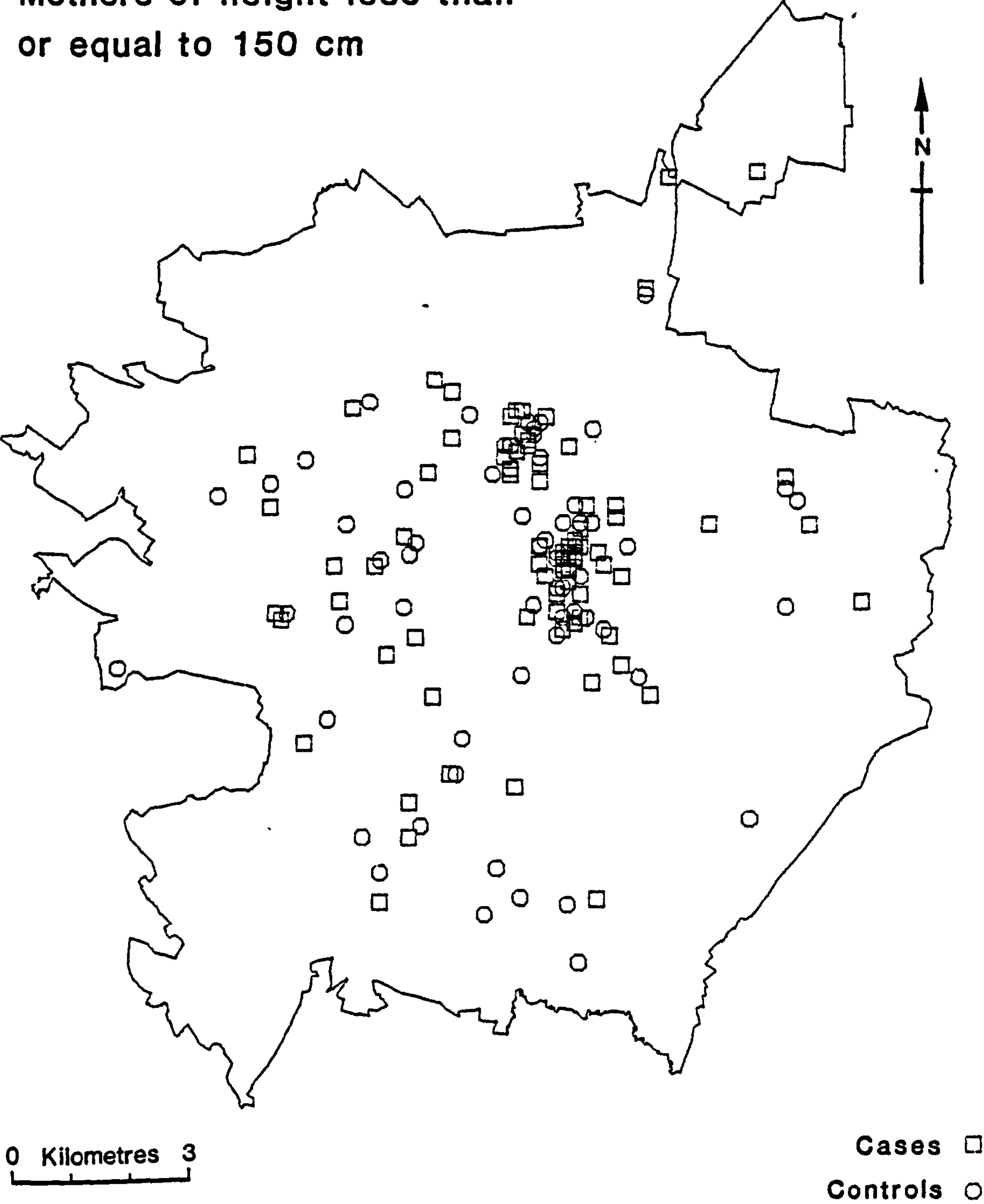


Figure 7.10

Zero parity mothers

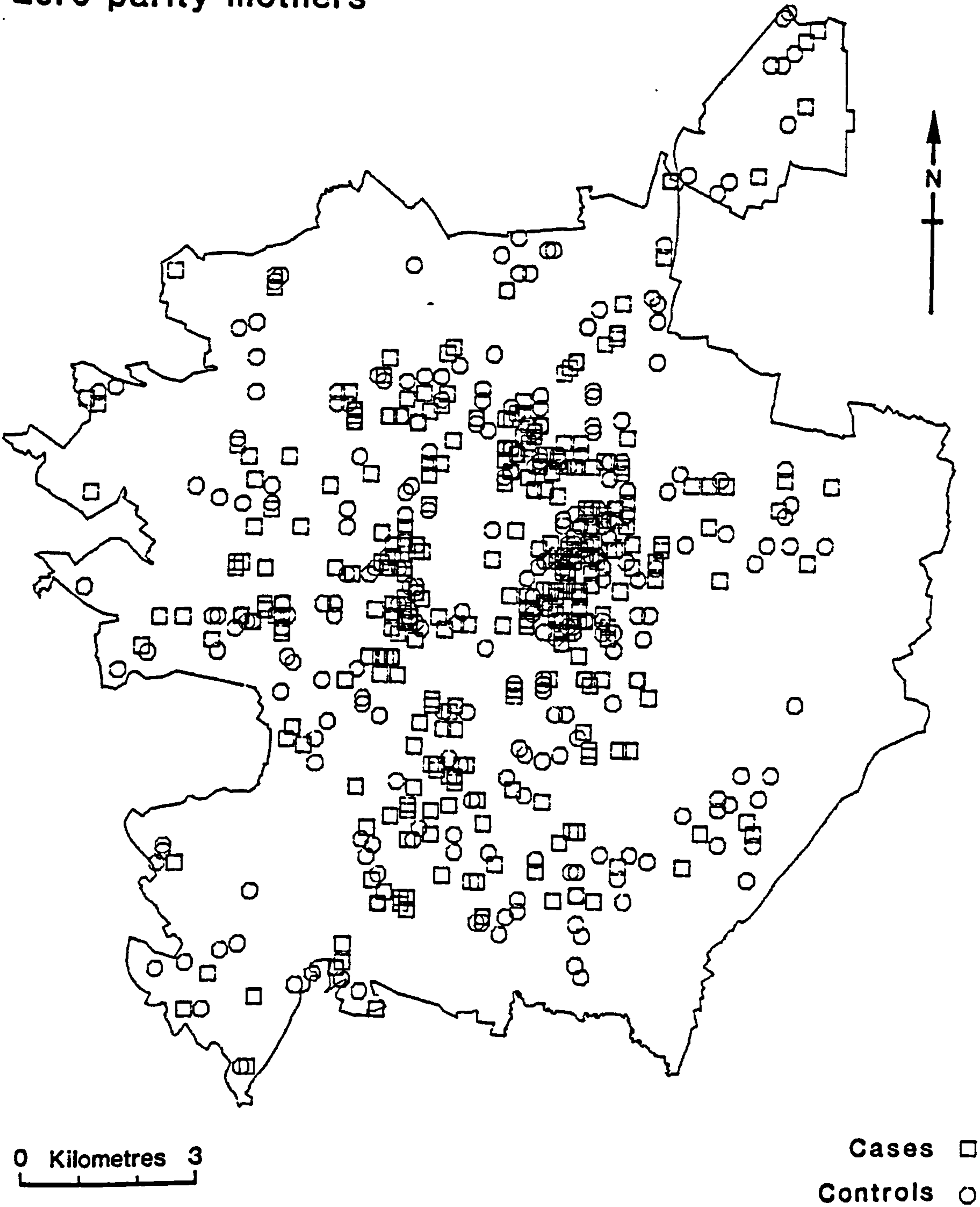


Figure 7.11

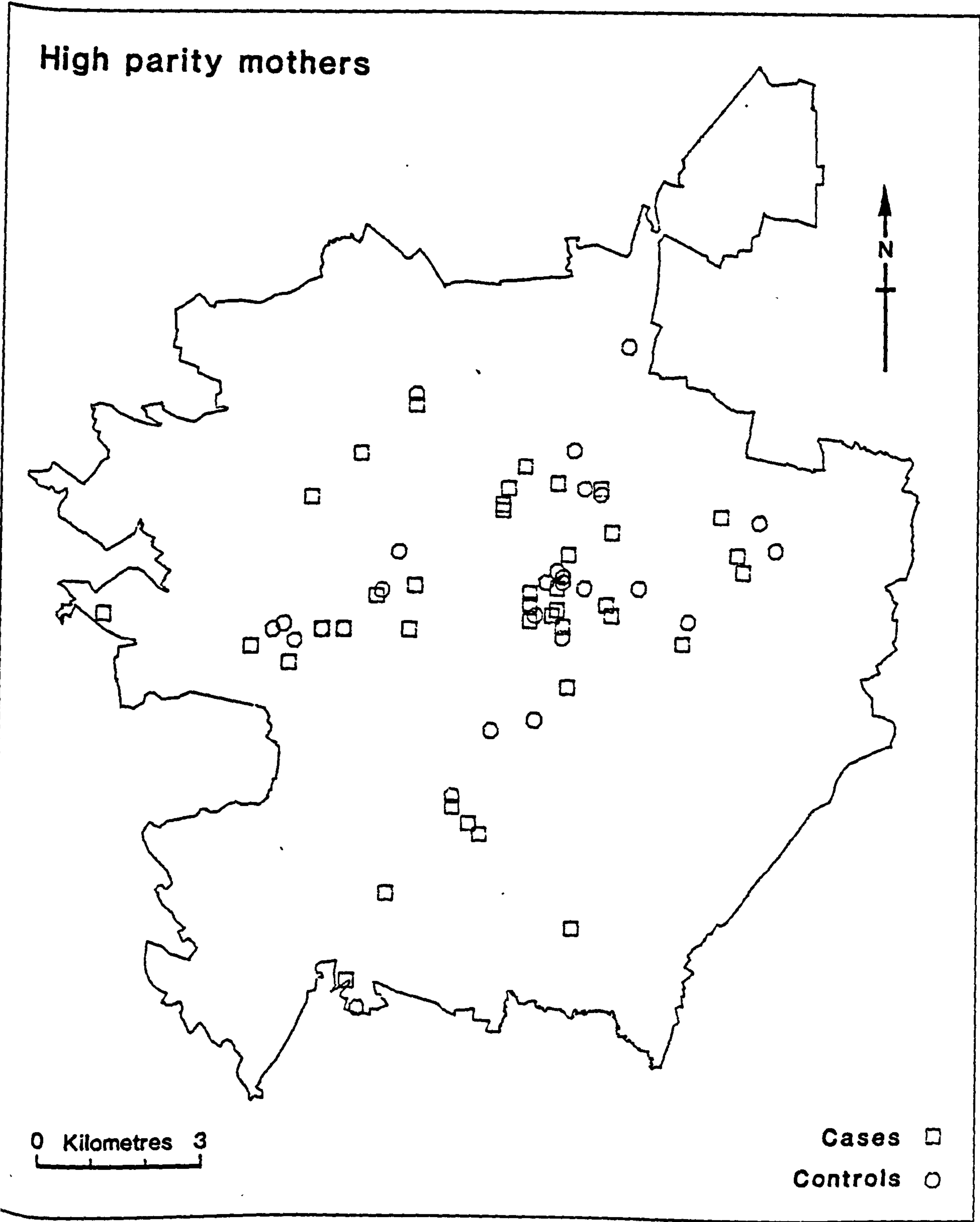


Figure 7.12

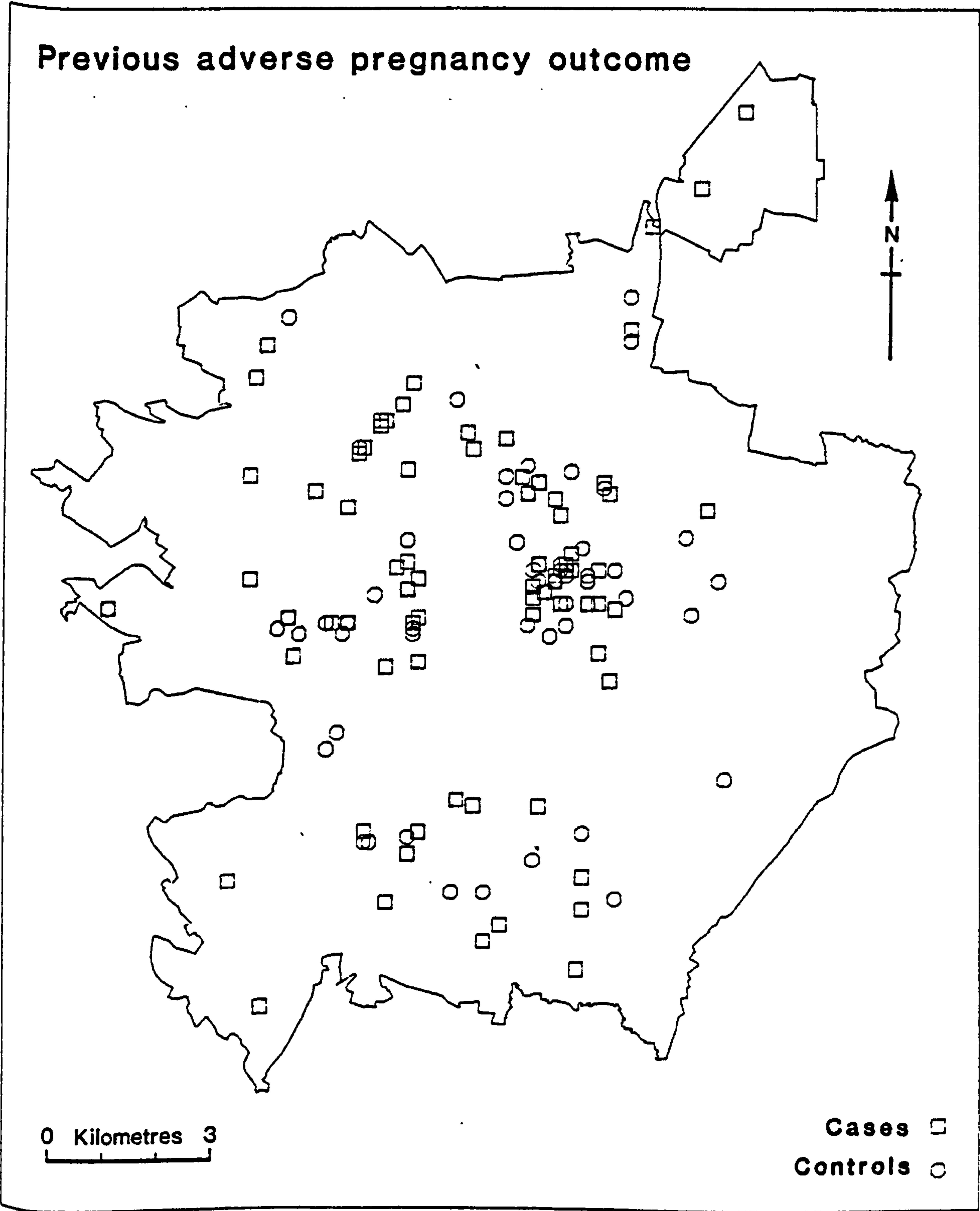


Figure 7.13

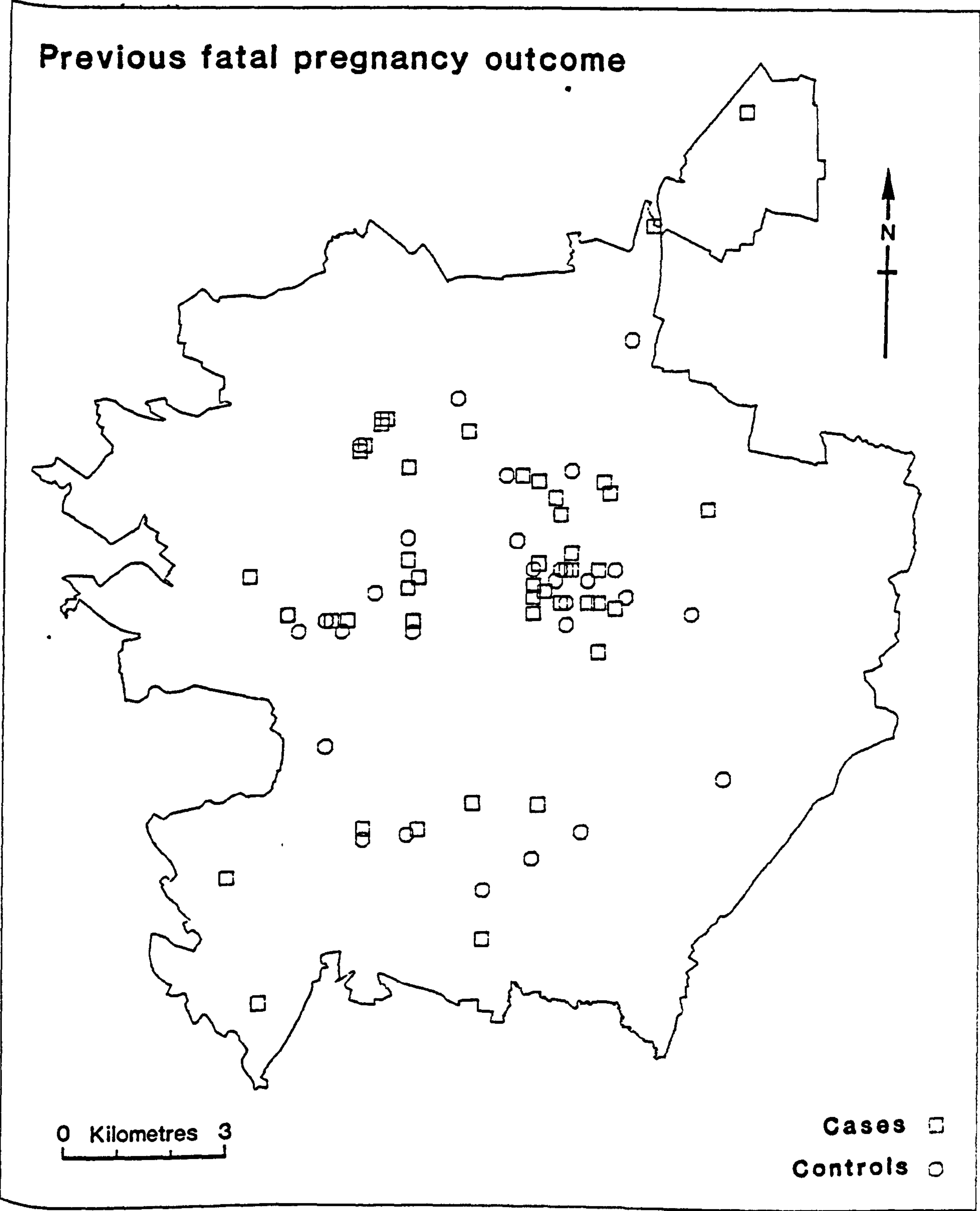


Figure 7.14

Mothers of social class IV or V

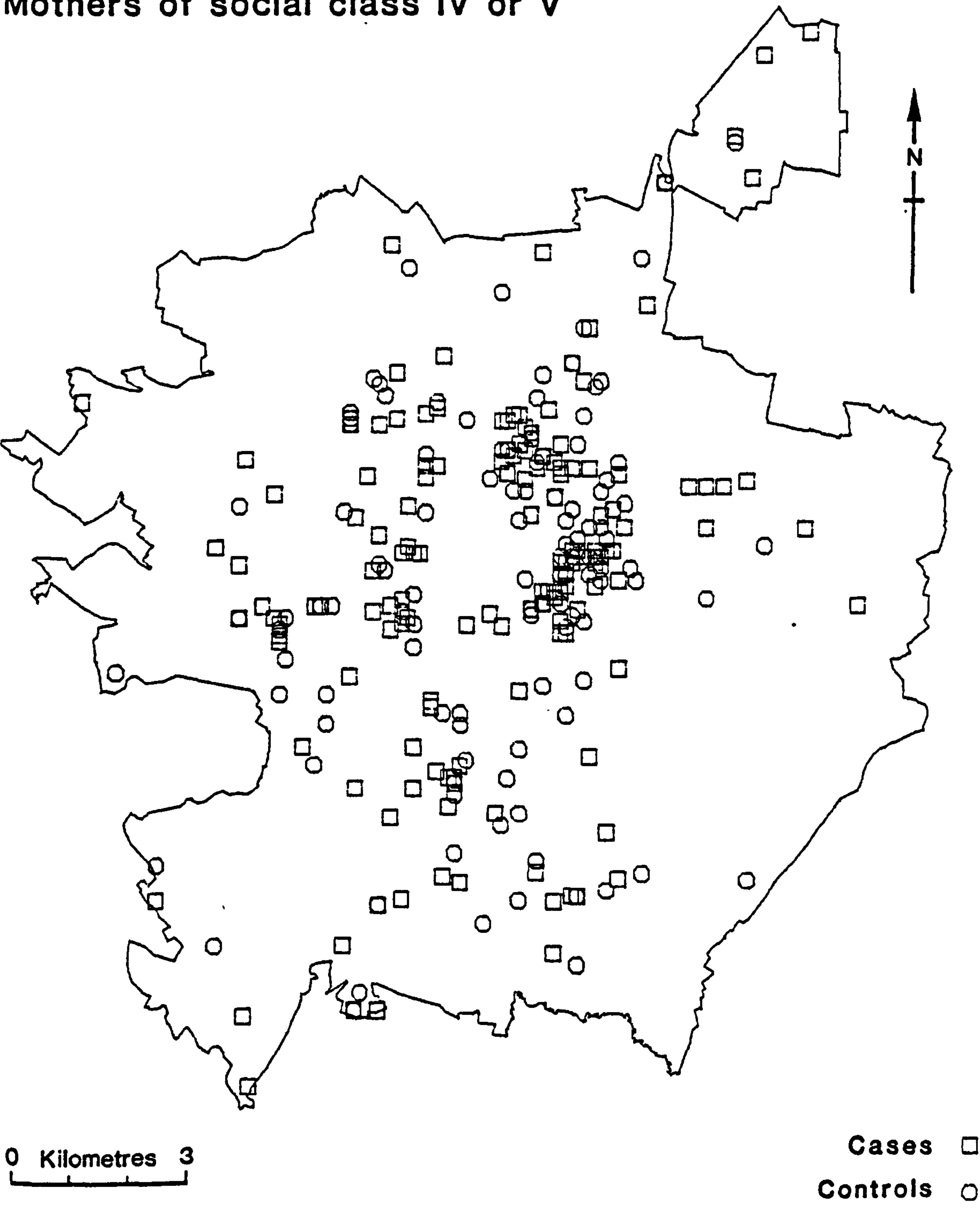


Figure 7.15

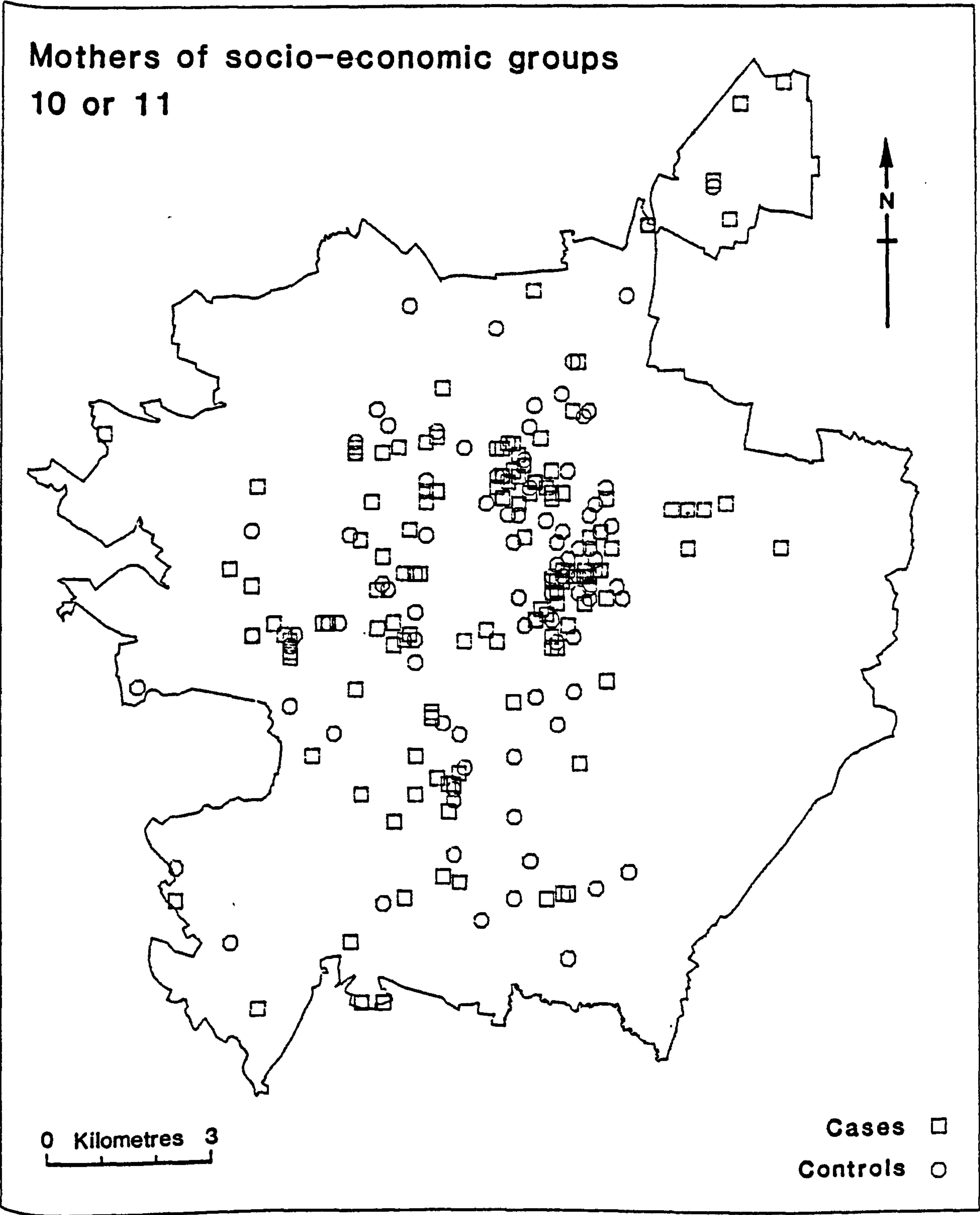


Figure 7.16

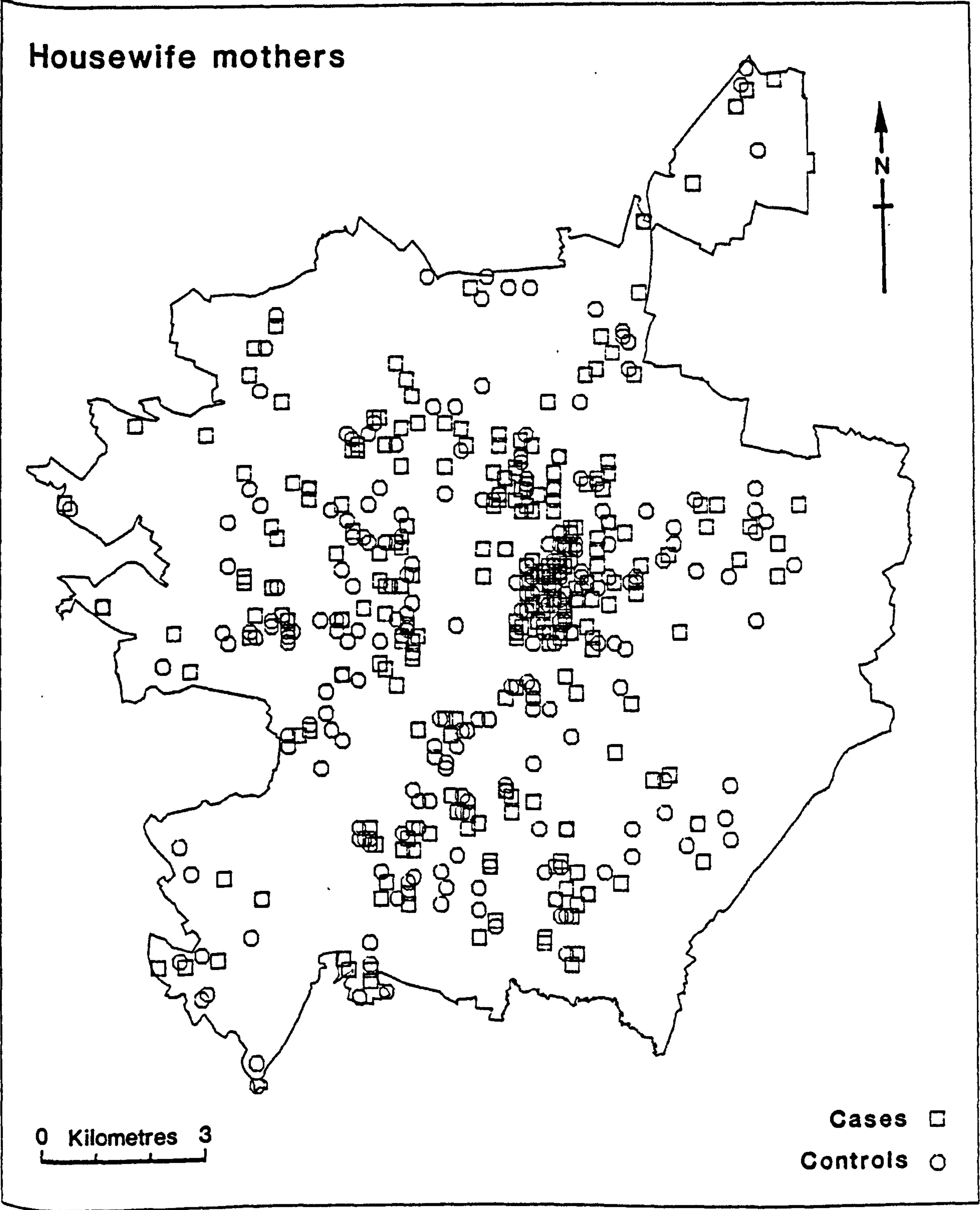


Figure 7.17

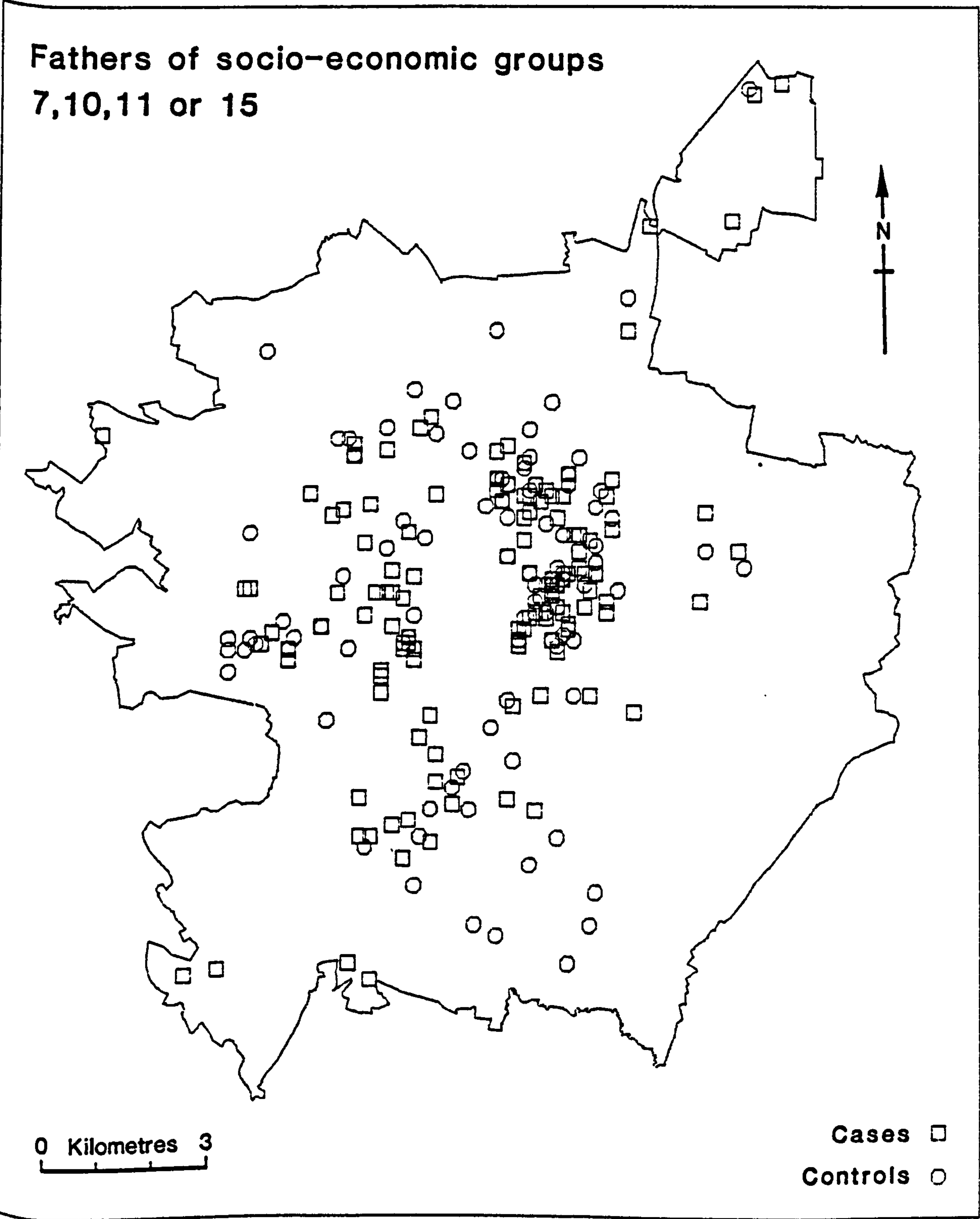


Figure 7.18

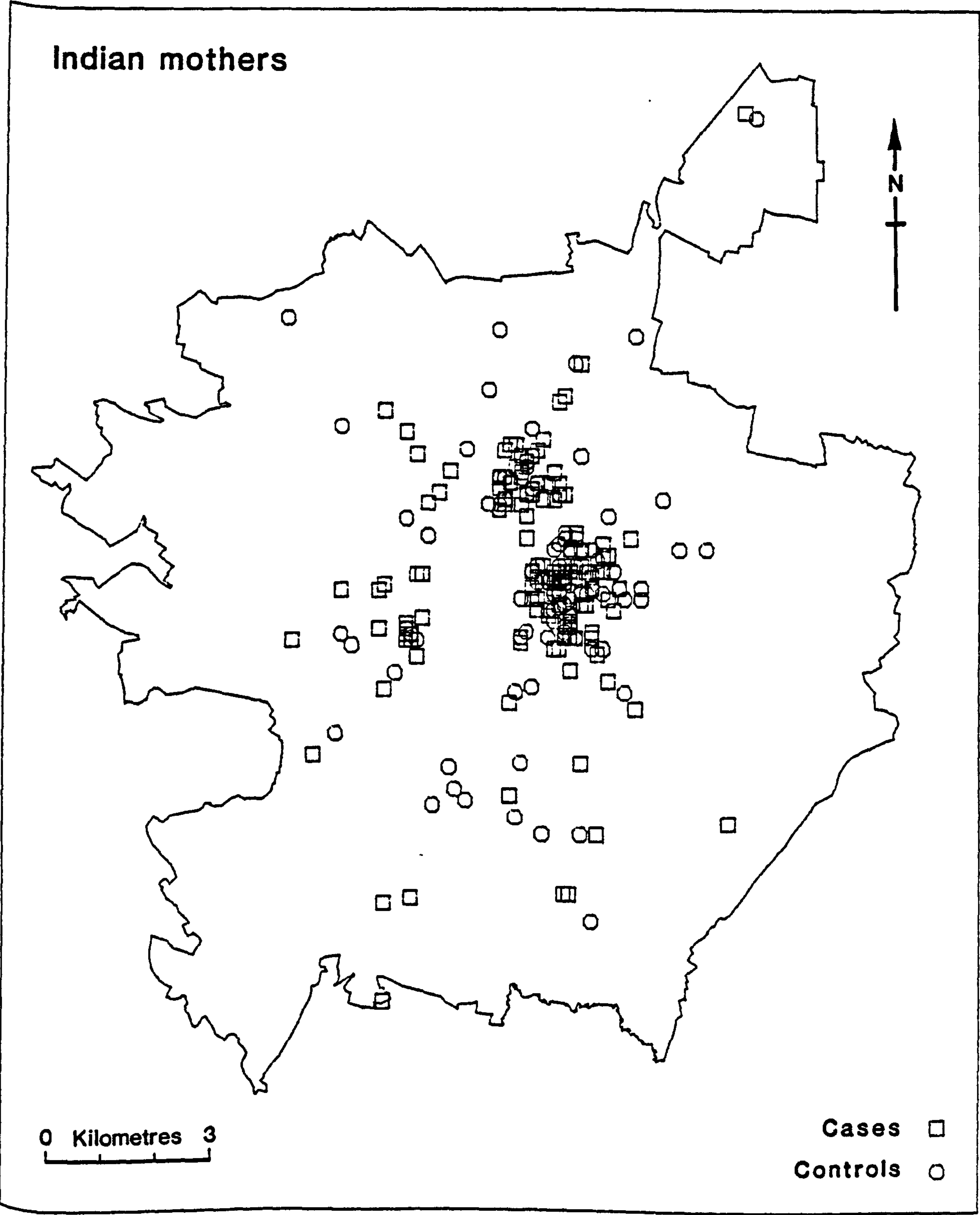


Figure 7.19

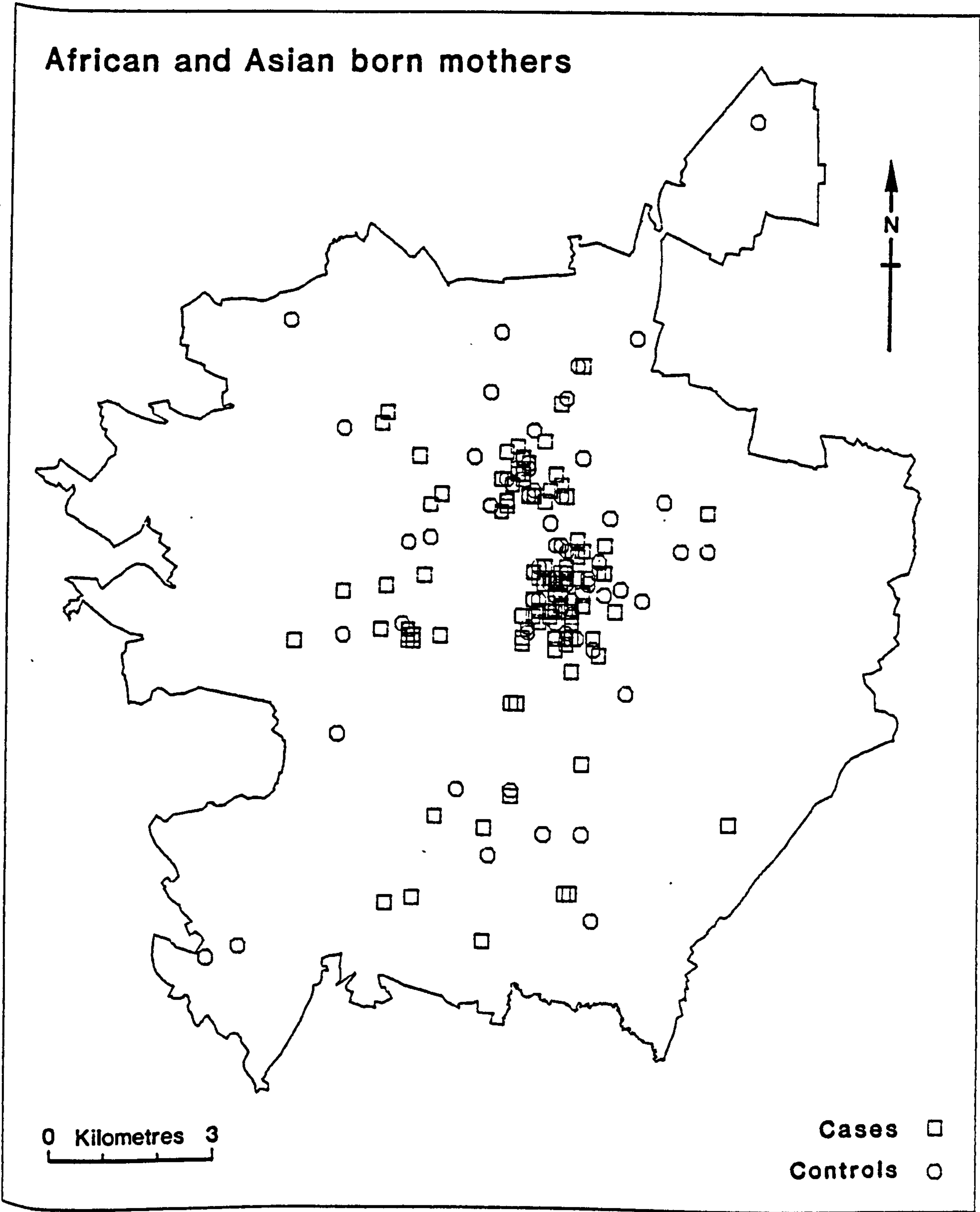


Figure 7.20

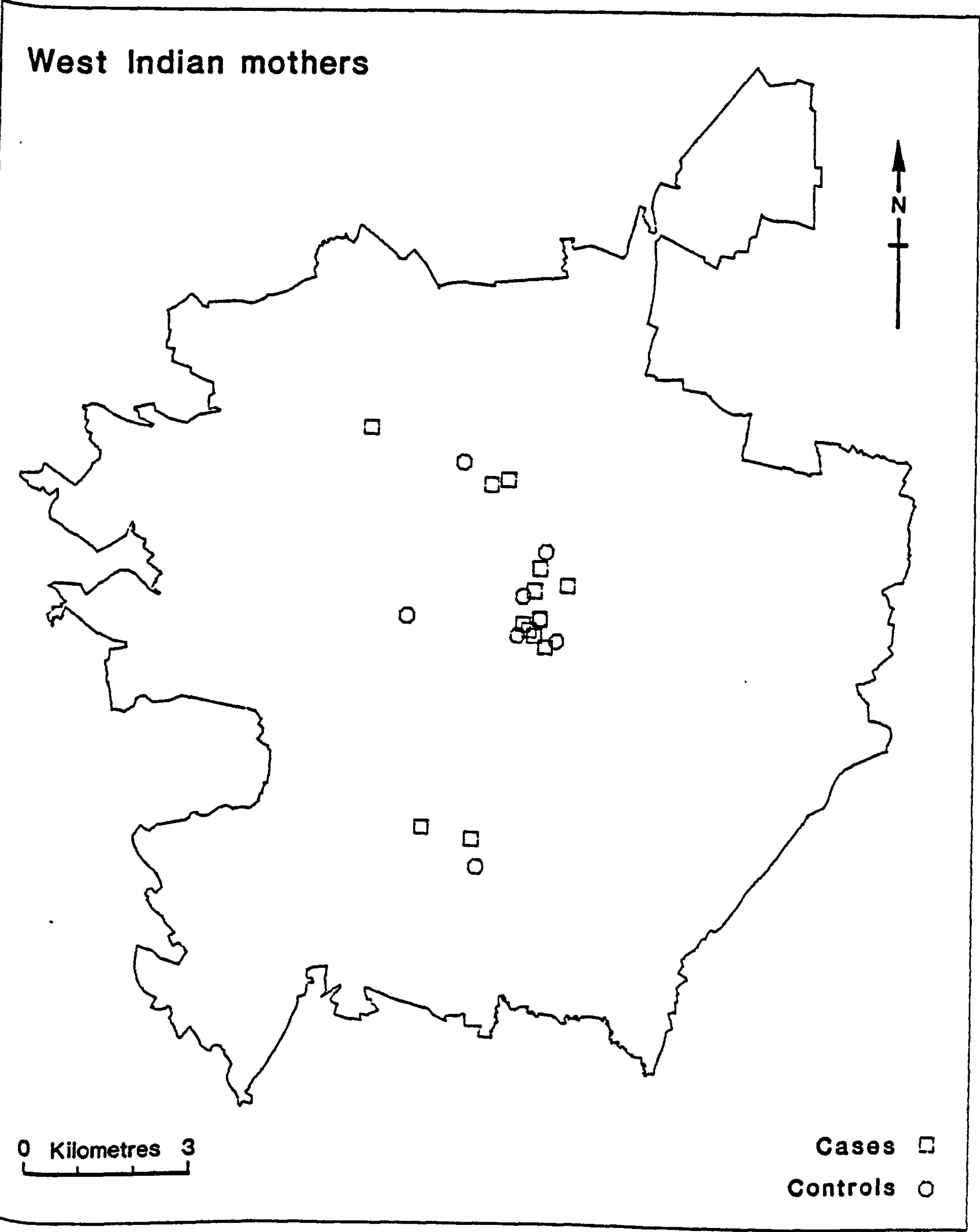


Figure 7.21

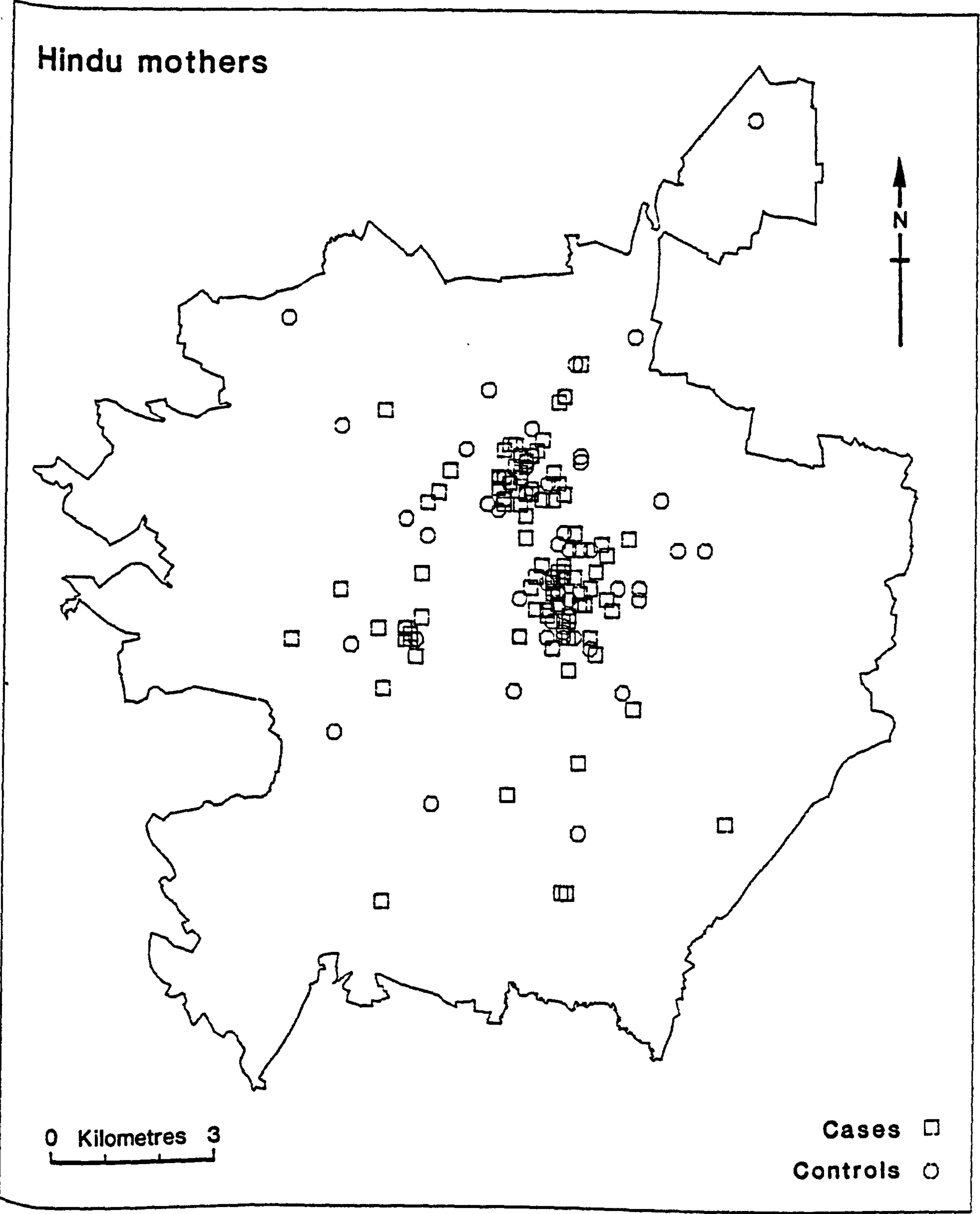


Figure 7.22

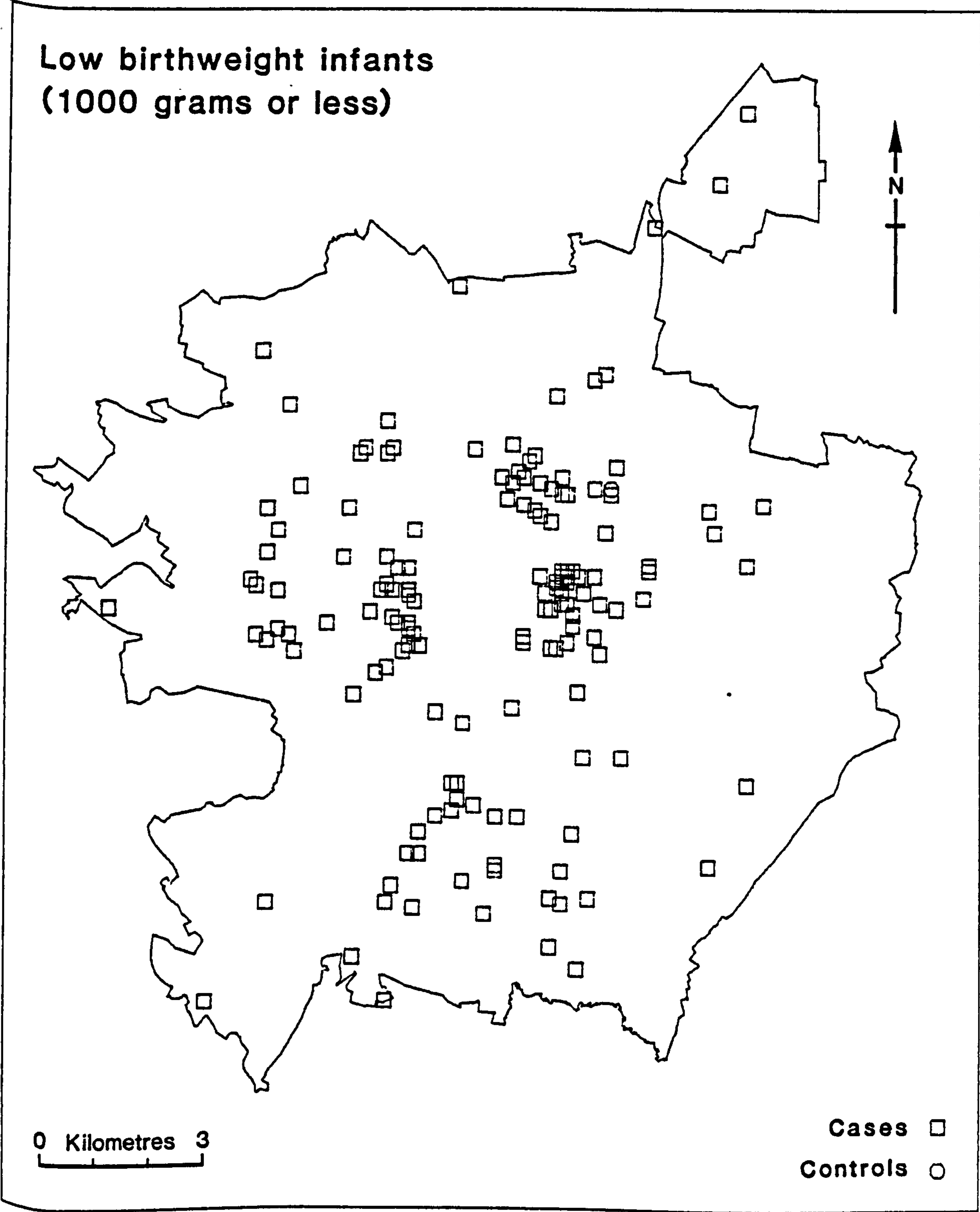


Figure 7.23

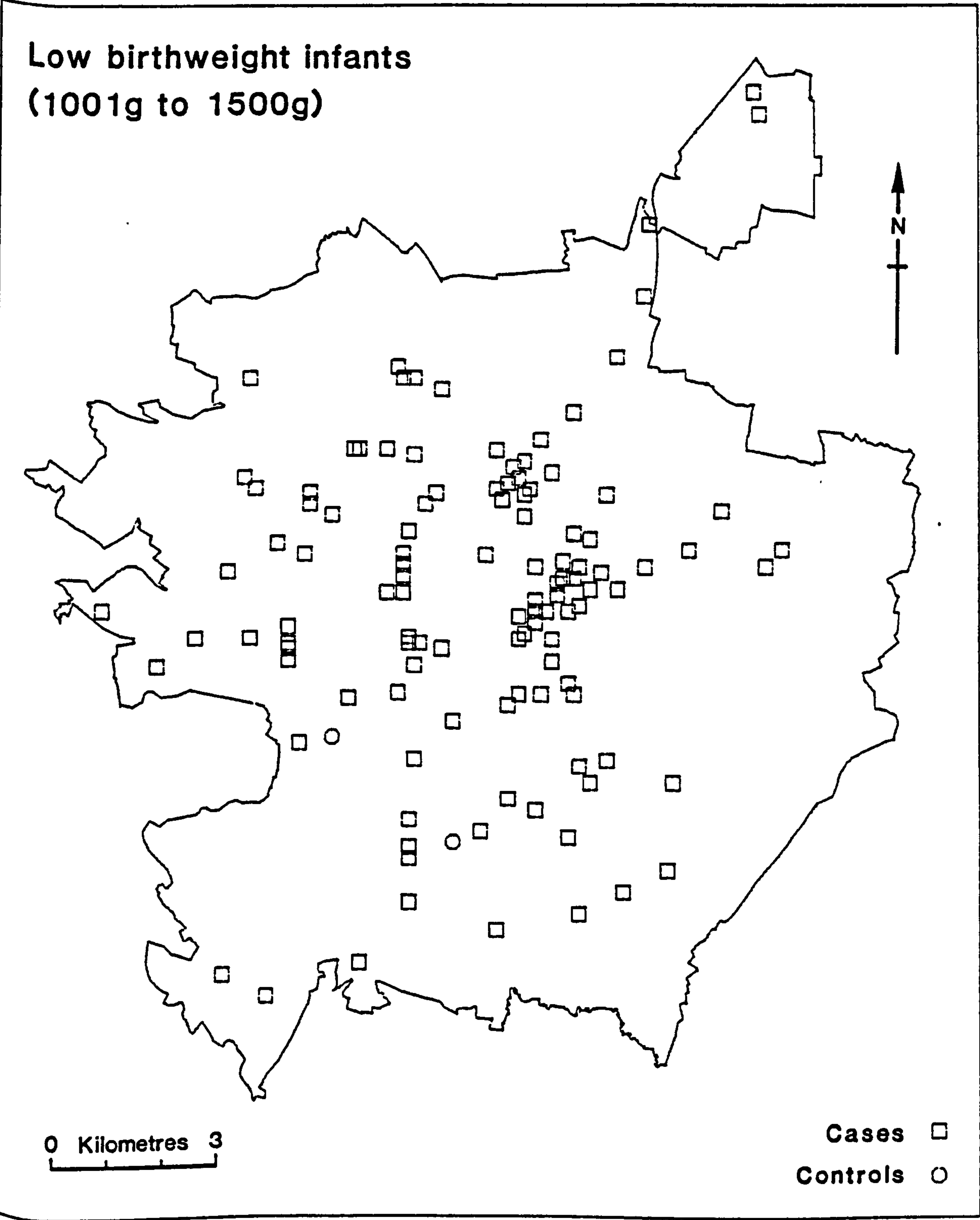


Figure 7.24

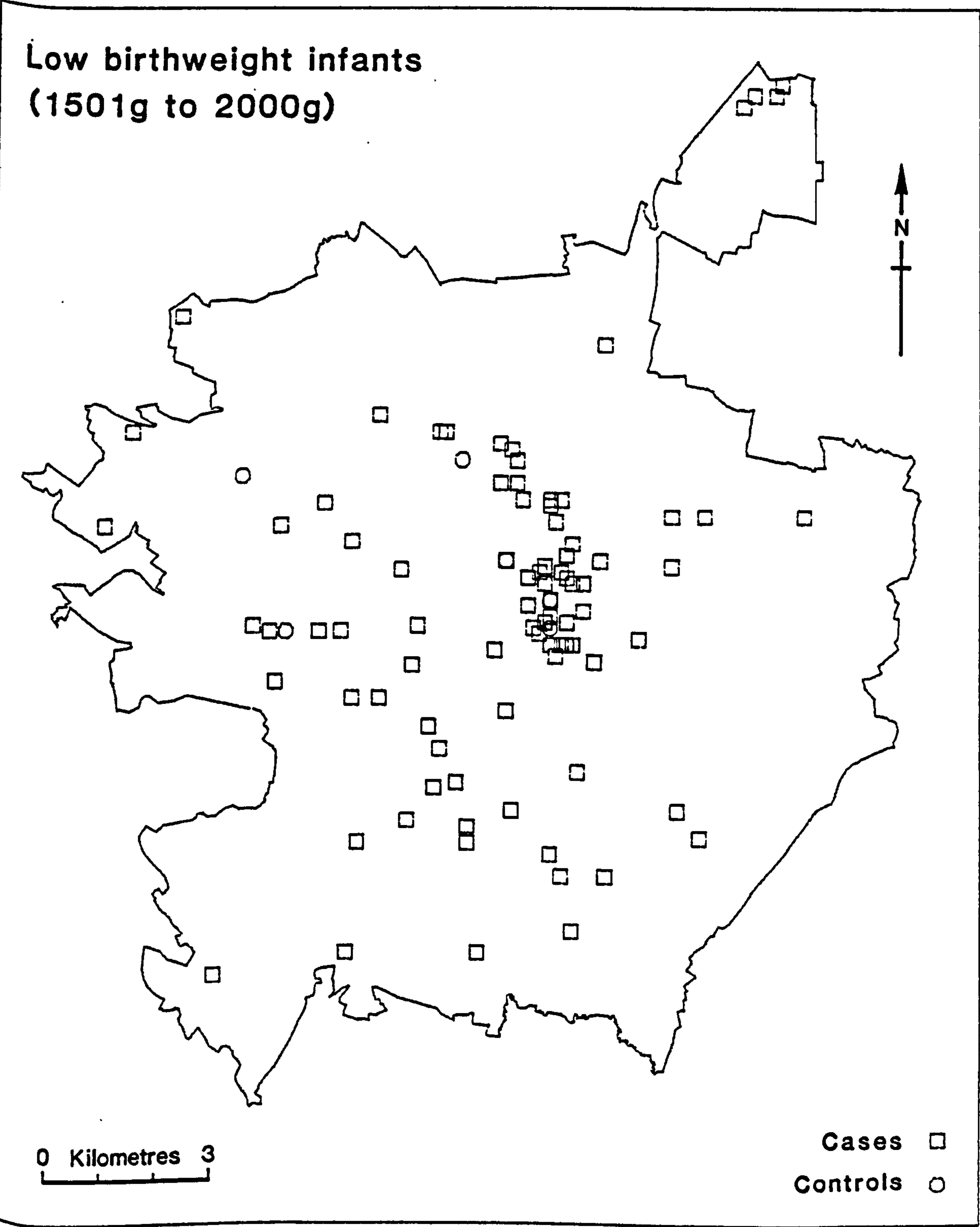


Figure 7.25

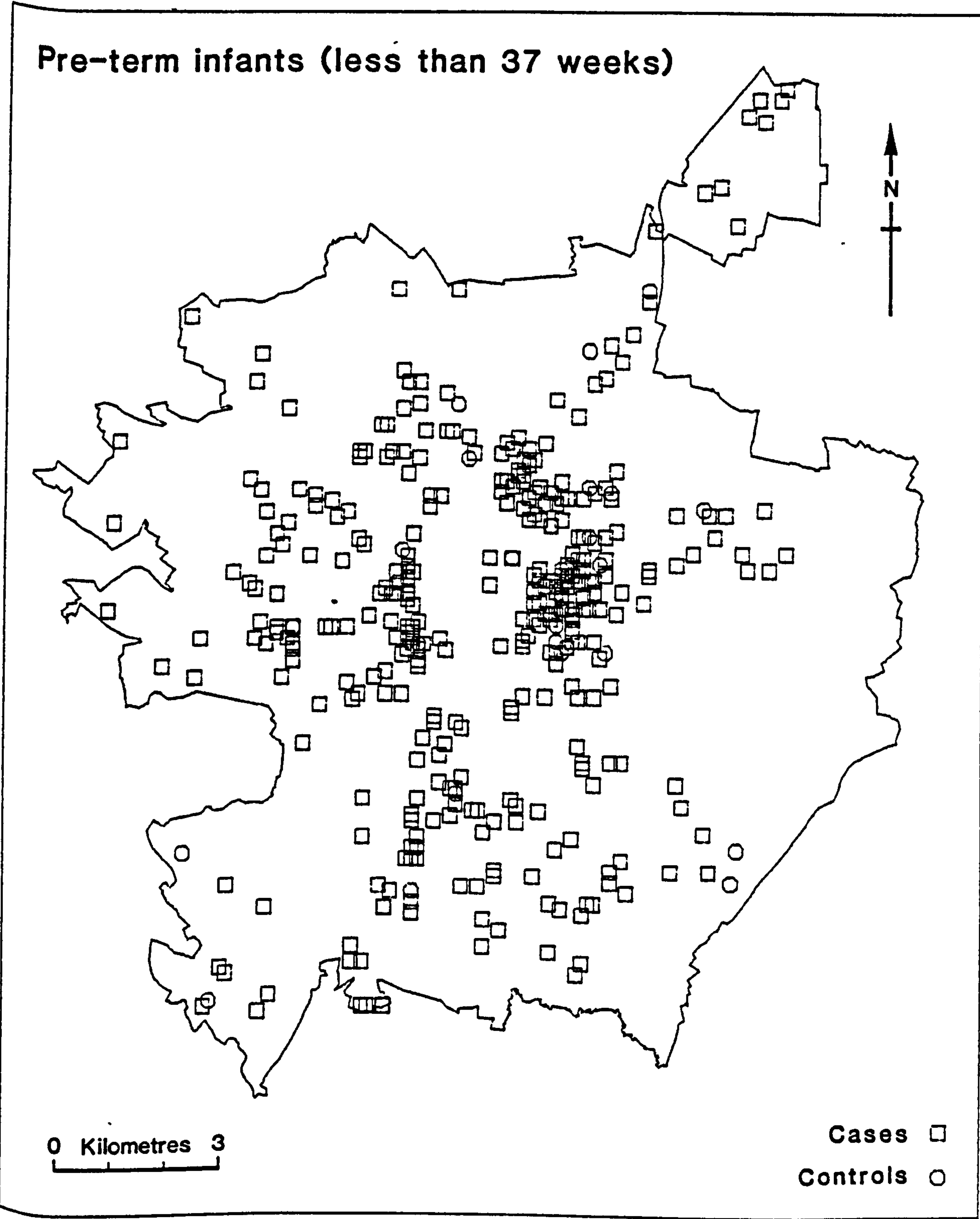


Figure 7.26

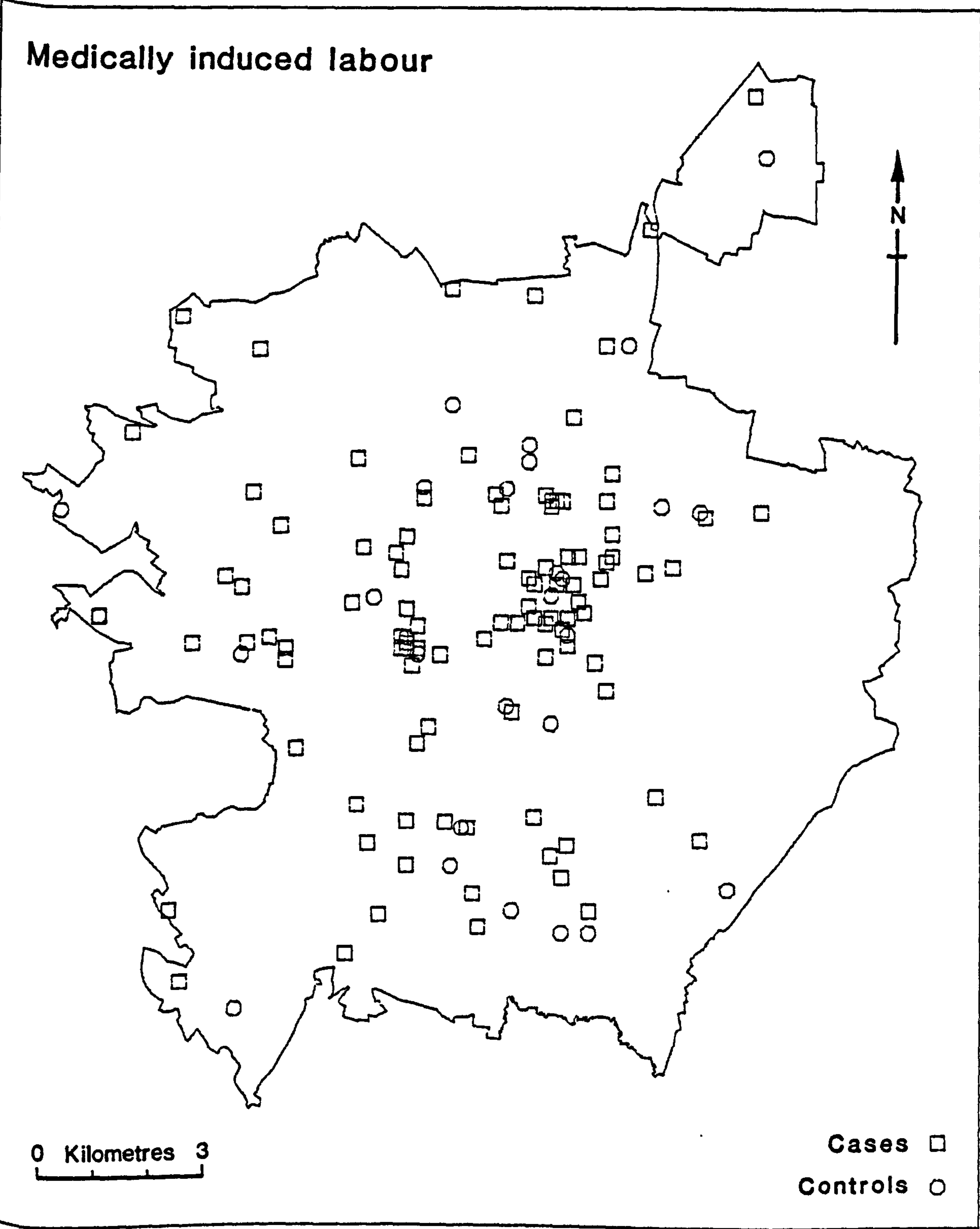


Figure 7.27

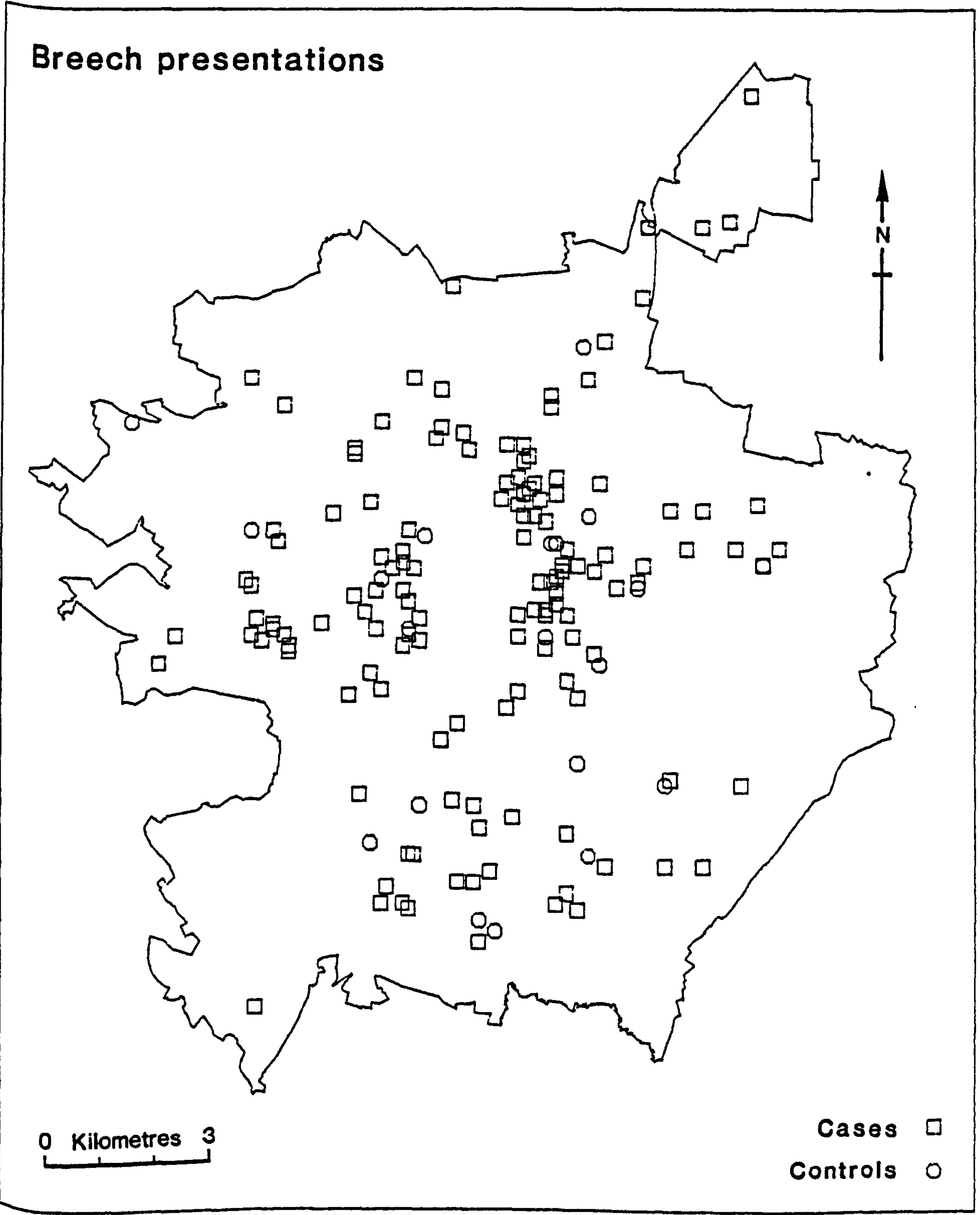


Figure 7.28

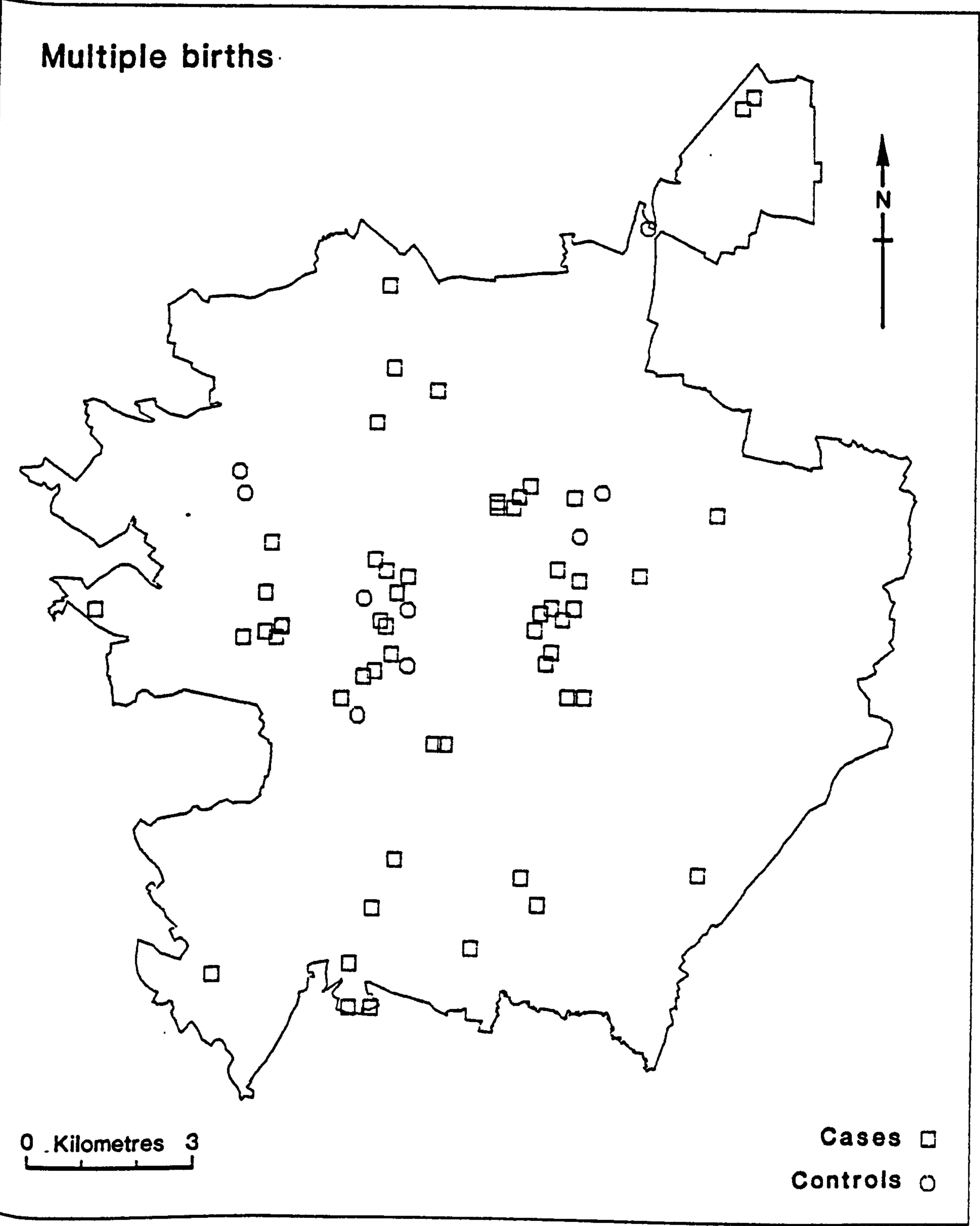


Figure 7.29

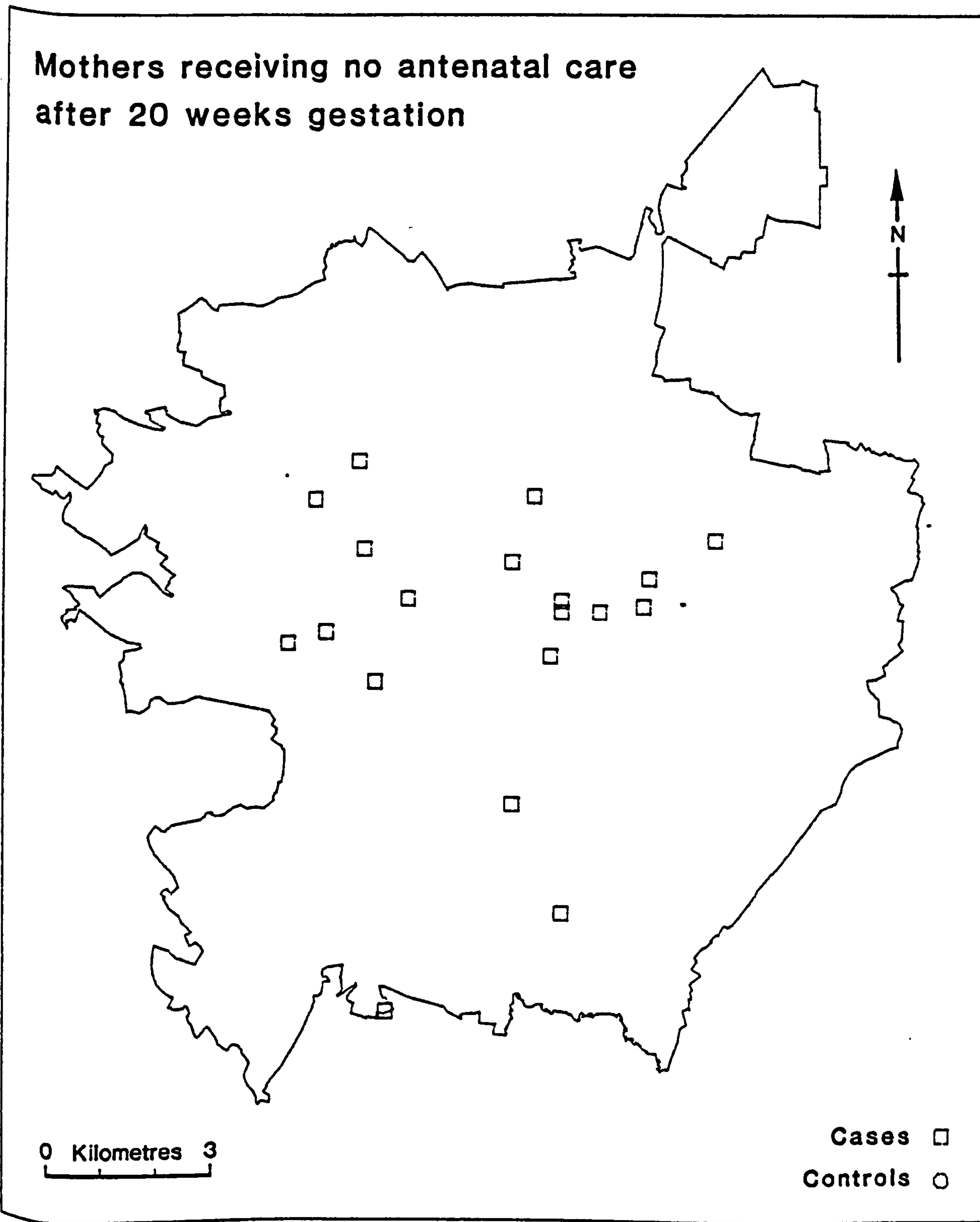


Figure 7.30

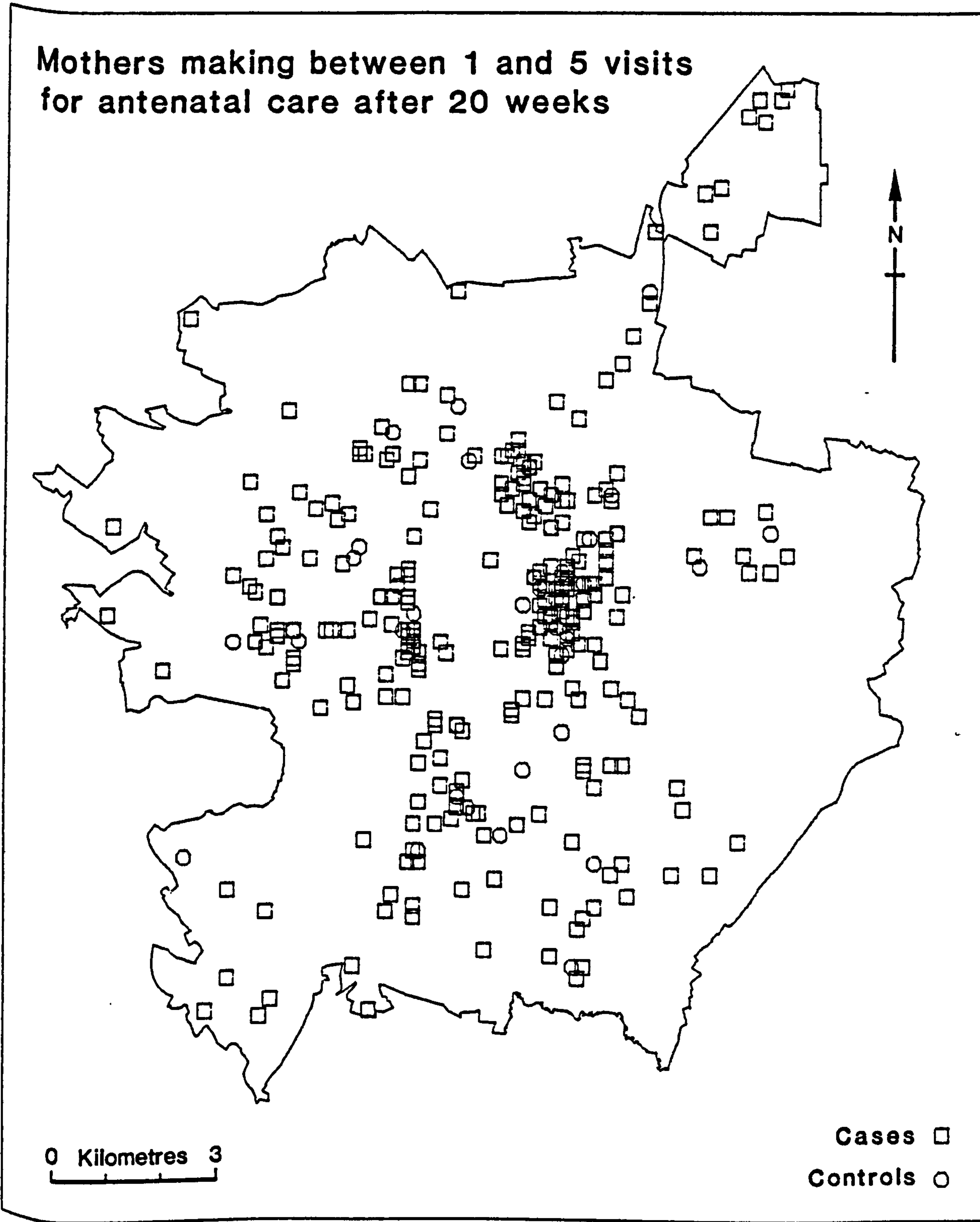


Figure 7.31

Mothers receiving more than 20 days
in-patient antenatal care

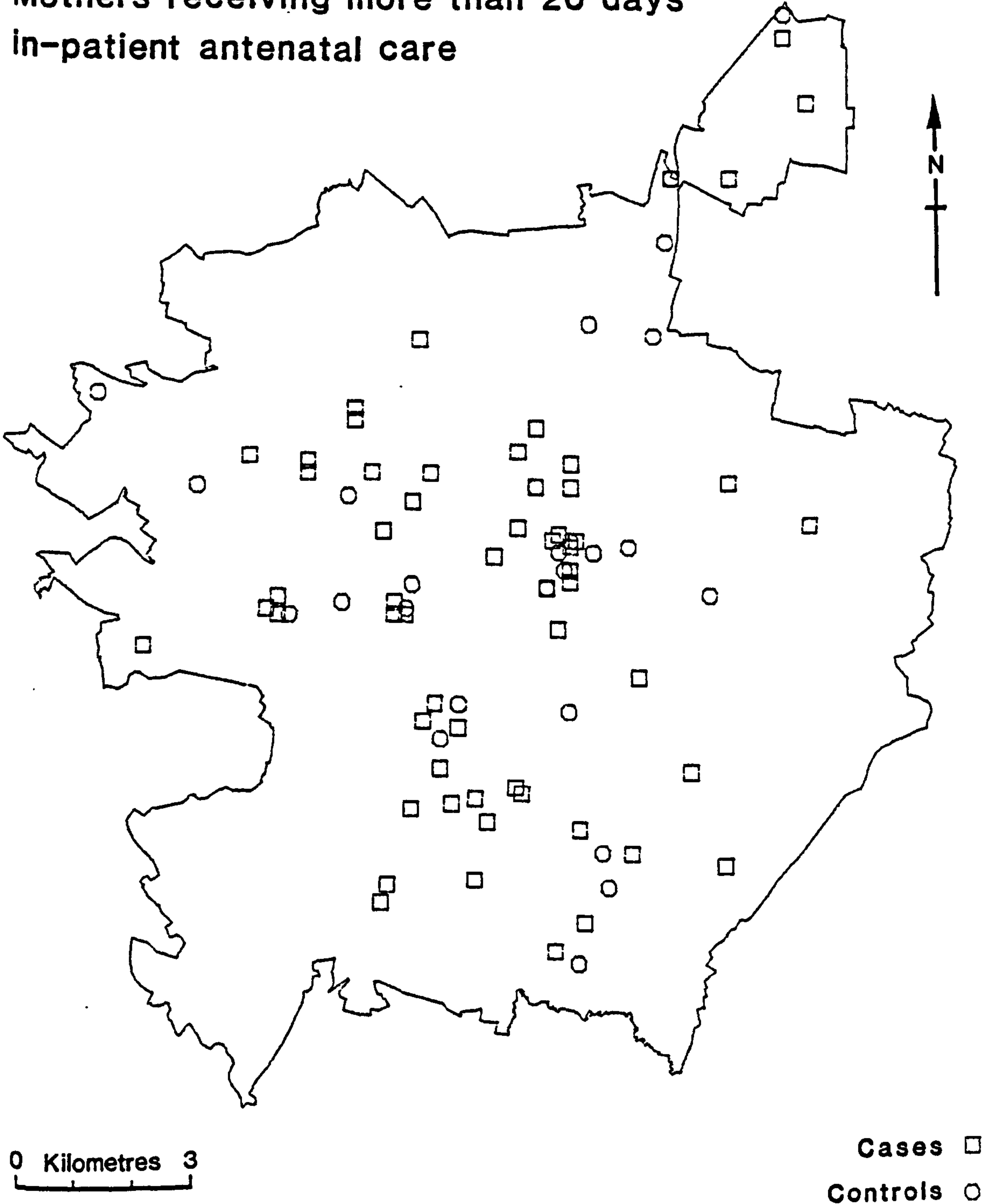


Figure 7.32

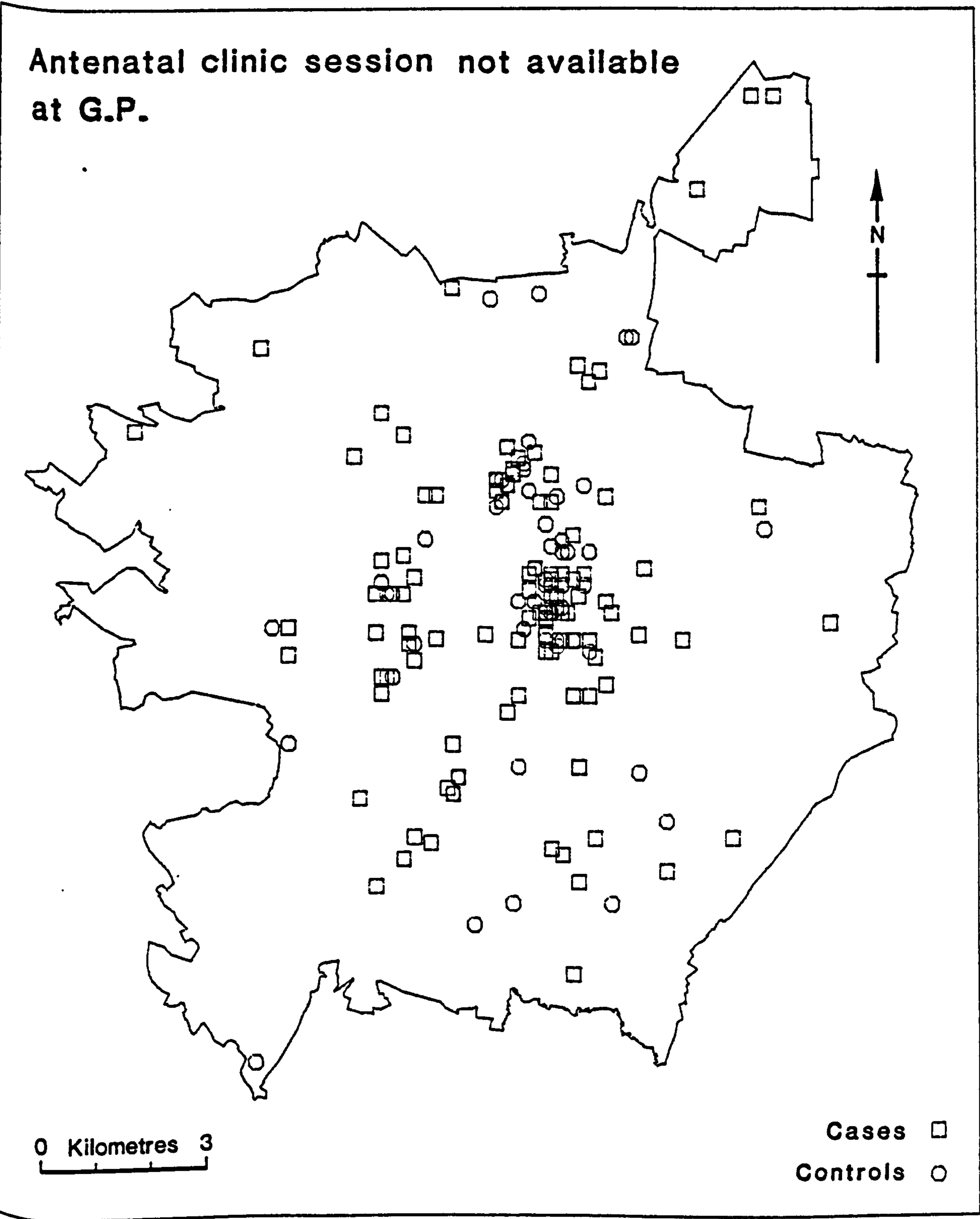
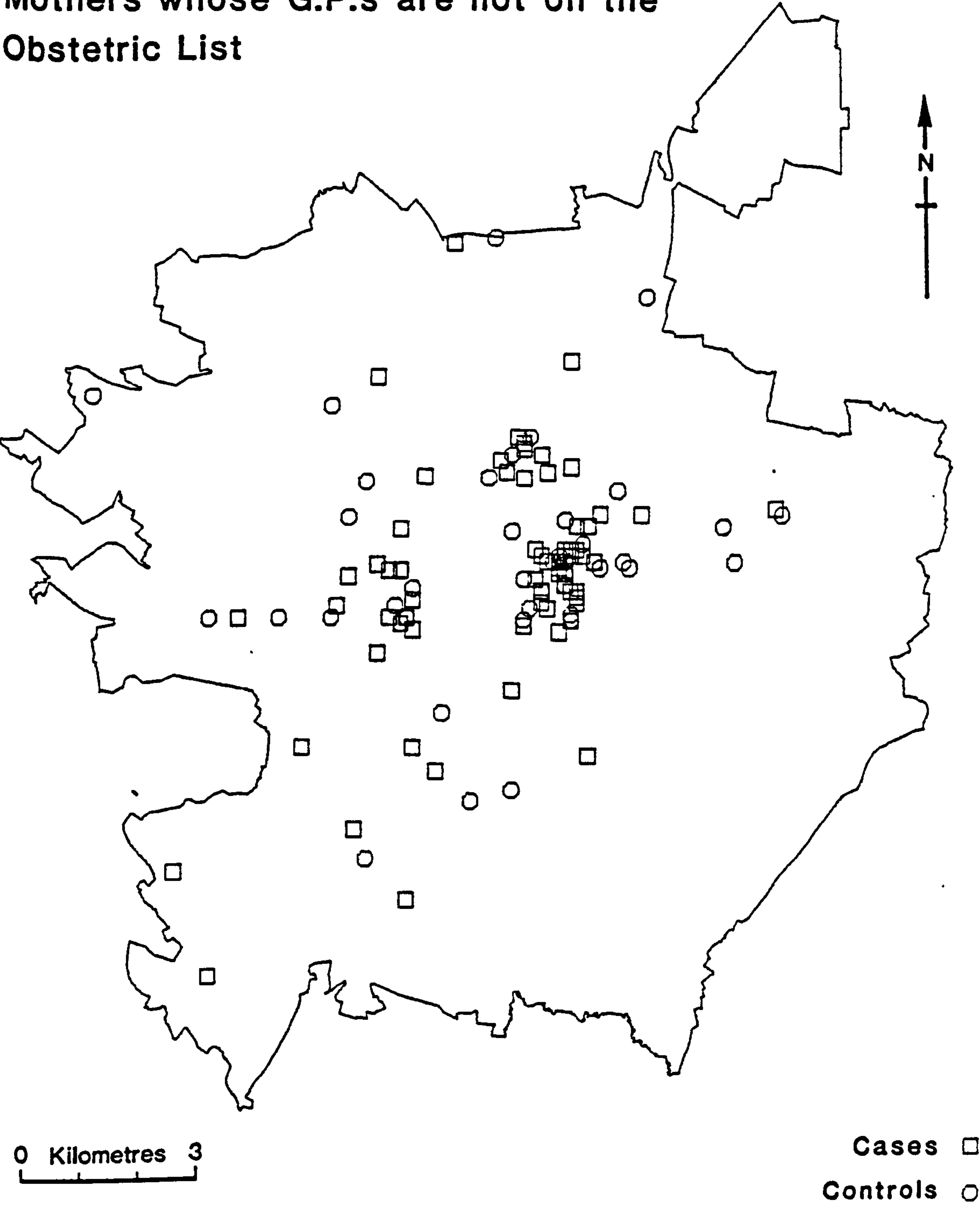


Figure 7.33

Mothers whose G.P.s are not on the
Obstetric List



Chapter 8. Further Data Analysis

8.1 Techniques for combining perinatal and census data: introduction

In previous chapters both census and perinatal data have been analysed in some detail to construct a broad picture of the social and biological environments into which the new baby is born. With the exception of certain variables which are common to both data sets, little attempt has been made to combine the two sets of results or even to assess the relative significance of the different variables.

The objectives of this chapter therefore, are to investigate alternative methods of combining perinatal mortality and census data, to examine further the relative importance of the different risk factors, and to establish whether any method or technique can provide new information for combatting the perinatal problem.

The joint analysis of census and perinatal data presents several problems, related both to the data themselves and to the techniques available for their analysis. The data values involved refer to individuals on the one hand (mothers and infants) and areas on the other (EDs or NSAs), a first problem is therefore deciding the most suitable means of comparing these. Several options are available, namely the comparison of absolute values, ratios, chi-square values or other derived statistics such as principal component scores. It is not essential for both data sets to take the same unit of measure. For example, it has been shown in Chapters 6 and 7 that the chi-square statistic is an appropriate value for displaying census data but, although valuable in identifying the highest risk areas, it lacks the sensitivity shown by the simple 'dot' type map in describing the detailed distribution of perinatal mortality (see Figs. 7.5 and 7.8). It was therefore shown to be most appropriate in Chapter 7

to compare the dot-type maps of perinatal mortality with the chi-square maps for census variables. This exercise was usefully extended by comparing subsets of the perinatal data, namely mothers or infants exhibiting risk characteristics, with the same census data, and could be extended still further by, for example, performing a principal components analysis on the latter and examining the distribution of component scores. The results of this and other procedures mentioned below will be described in the next section.

A second, and related problem, is the need to standardise for the perinatal data. To this end, both ratio and chi-square values were calculated in section 7.4.3, each by comparing the distribution of mortalities with that of births. The effects of this standardisation on the calculation of correlation coefficients for areal association between mortalities and census variables are shown below.

A third feature of the perinatal data which causes problems in further analysis is the distribution and low density of observations. In addition to the danger of producing low and non-meaningful rates for statistical purposes, any aggregation into areal units will result in areas with zero observations, too many of which are unsatisfactory visually and unhelpful in further cartographic analysis. This is a particular problem in, for example, the calculation and plotting of a 'risk surface' to show areas of high and low risk of perinatal death.

Finally, neither perinatal nor census data in their present format fulfil all the assumptions made by the more sophisticated statistical techniques. Because of this it is doubtful that there exists a statistical theory exactly suitable for these data. If however they were reduced to fewer, but larger, classes it might

be possible to model the relationships between variables using for example, log-linear methods.

8.2 Some experiments in combining the data sets

8.2.1 Principal components analysis

The first exercise in analysing further the two data sets was to perform a principal components analysis (PCA) on the census data. PCA is one of a family of data reduction techniques, used essentially as inductive tools for searching out the basic dimensions of a data set (Clark et al., 1974). In the current project the 'dependant' variable, perinatal mortality, has already been compared with individual 'independant' variables from the census; the objective of this PCA is to reduce the number of independant variables and enable a search for similarities between perinatal mortality and the new dimensions. Here the PCA is basically an exploratory or descriptive exercise, the main advantage being that it allows the simultaneous consideration of a large number of variables.

A data matrix of order $n \times k$, where n refers to variables (for example, socio-economic group) and k refers to cases (NSAs), was created. The matrix was then transformed and the data values converted to standard scores before calculating correlation coefficients between pairs of variables ('R-mode' analysis). In the absence of unreasonably high correlations between pairs of variables the component matrix was then calculated. The advantage of performing a PCA rather than a factor analysis is that it produces a unique orthogonal solution in which successive components each maximise the variance explained until the final component has accounted for the total variance. Rotation procedures are not therefore helpful and were not applied. Table 8.1 summarises the results of the PCA. Figures 8.1 to 8.7 show the distributions of each of the components. CHORO's 'percentile'

Table 8.1 Principal Components Analysis - results

Number of variables (n) = 34
Number of areas (k) = 227

Component	Eigenvalue	% of variance	Cumulative % of variance
1	8.65	25.4	25.4
2	5.21	15.3	40.8
3	4.46	13.1	53.9
4	2.44	7.2	61.0
5	1.79	5.3	66.3
6	1.64	4.8	71.1
7	1.22	3.6	74.7
8	0.99	2.9	77.6

Variable loadings on components with eigenvalues greater than 1.0
(highest 10 variables only)

Component 1

MNOCAR	.917
EMZACT	- .854
HSOWN	- .774
HHSING	.748
EMEACT	- .713
DMIDWK	- .704
HSCOUN	.697
DOLPNS	.664
DYNPNS	.655
EMUNEM	.635

Component 2

HSCRWD	.715
DYNGWK	.699
EMUNEM	.642
DYNPNS	- .583
HHCHIL	.580
ETNCWP	.562
MMIG	.511
HSAMEN	.498
DYNGCH	.483
DOLPNS	- .479

Component 3

HSRENT	.678
HSSHAR	.664
DOLDCH	- .659
SEEDUC	.634
MMIG	.623
HHSING	.559
HHXFAM	- .503
HHCHIL	- .481
SENMAN	.472
SESMAN	- .404

Component 4

EMSEMP	.518
EMSTU	.518
HHXFAM	.443
EMEACT	- .427
HSFLAT	- .411
HSCRWD	.377
EMWOM	- .372
HHLPAR	- .342
HSOWN	.339
DMIDWK	- .331

Table 8.1 (cont.)

Component 5

DOLDWK	.585
HSAMEN	- .432
DYNGCH	- .404
SESMAN	- .337
HSCOUN	.316
EMUNYG	.286
SEEDUC	.281
HSOWN	- .279
SEPROF	.265
HSRENT	- .265

Component 7

SESS	.570
ETNCWP	.522
DOLDCH	- .337
HHCHIL	- .312
EMMWOM	.221
HSFLAT	.213
HSOWN	.190
EM2ACT	.173
DOLDWK	.171
HSCOUN	- .170

Component 6

DOLDWK	.464
SESMAN	.449
HSFLAT	- .425
HSAMEN	.399
HSSHAR	.363
SENMAN	- .360
ETNCWP	- .294
HSRENT	.284
DYNGCH	- .282
EMMWOM	.226

shading option was selected in order that areas with the highest positive and negative component scores could be most clearly identified.

In the following paragraphs labels are given to indicate the major dimensions of each component, these labels should not be seen as definitive since in most cases the variable weightings are too weak to justify this.

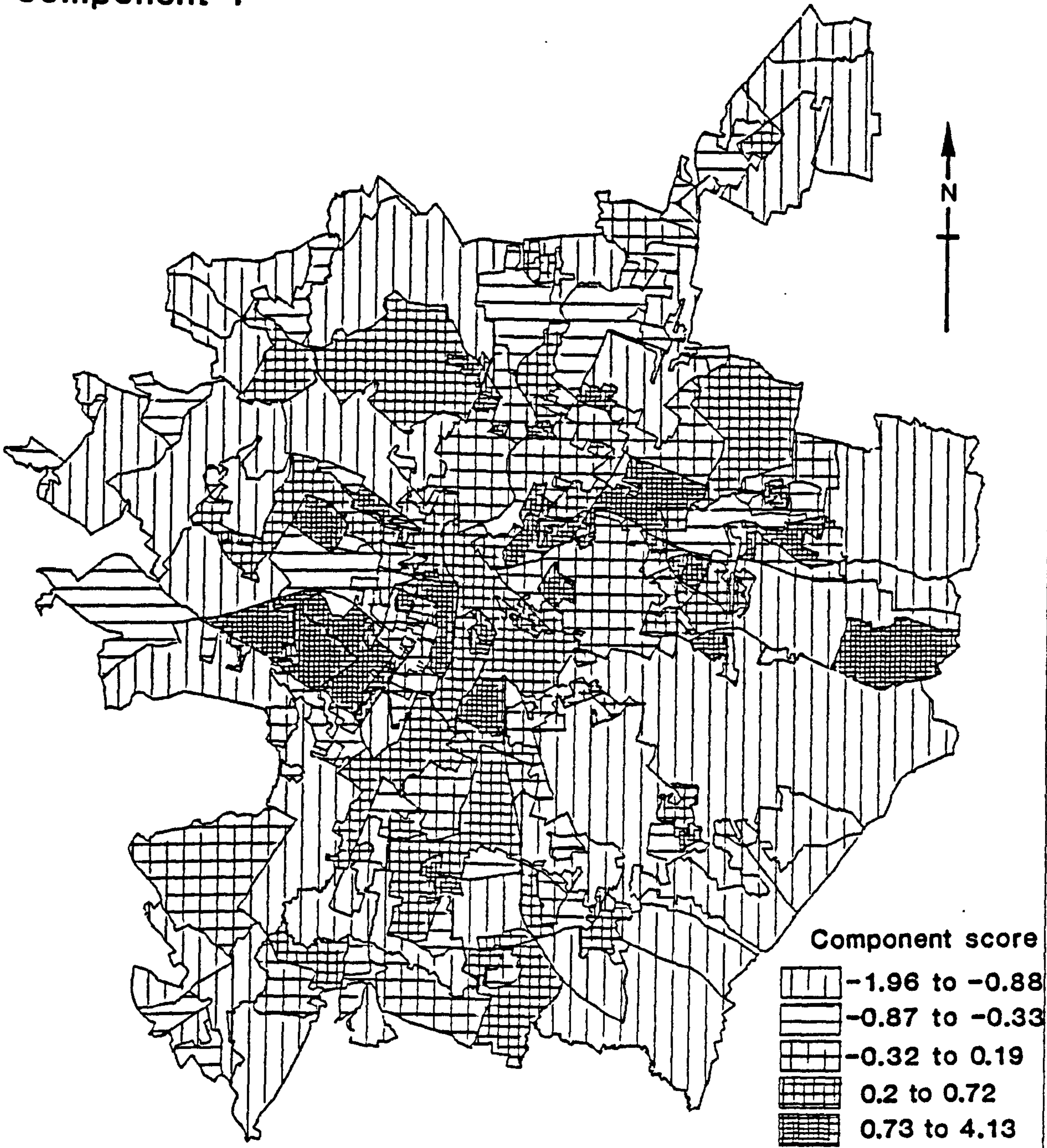
Although not significant among the component loadings, the distribution of professional and managerial households (see Figure 6.22) appears to be the inverse of component 1 (Figure 8.1). In the main this component identifies the work status of the resident population - suggested by high positive loadings on zero car ownership, single person households and people of pensionable age, and by high negative loadings on the economically active population and the middle working age group. A negative loading on owner-occupation and a positive loading on council housing complete the description.

The loadings on component 2 describe the life cycle distribution within urban areas. Young people, overcrowded conditions, unemployment, large households, ethnicity and migrancy all have high positive loadings whilst the pensionable age groups are the only variables with negative loadings. Not surprisingly, Figure 8.2 most closely compares with Figure 6.4 showing persons aged 16-24.

Component 3 has the highest weightings from tenure variables, specifically privately rented and shared accommodation. In addition, the better educated, one year migrants and single people also have high positive loadings, whilst variables indicating family life - such as older children and mixed families - are represented by high negative loadings. These characteristics

Figure 8.1

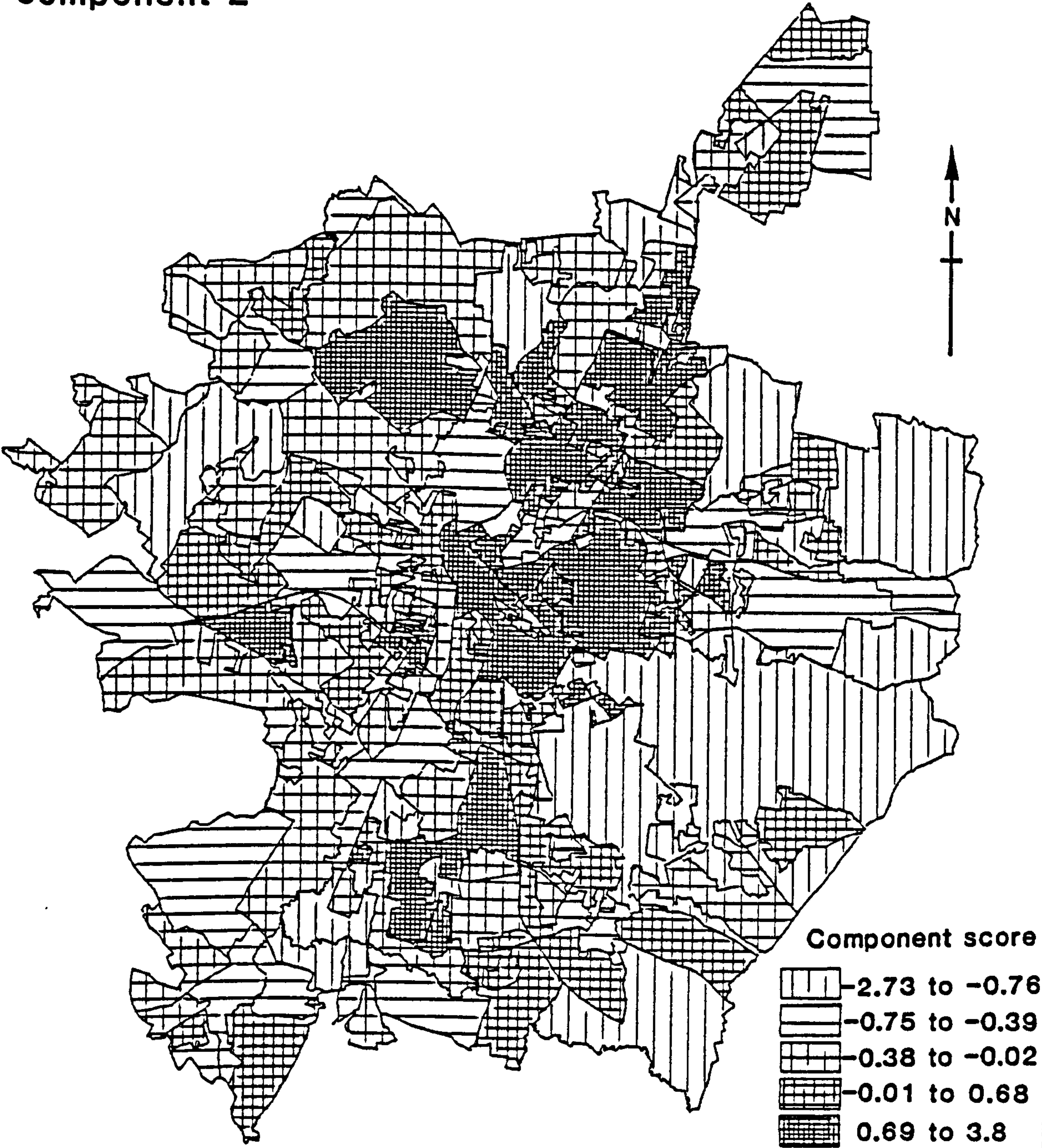
Component 1



PLOTTED BY >> CHORO << S4/06/07.
LEICESTER UNIVERSITY

Figure 8.2

Component 2



0 Kilometres 3

PLOTTED BY >> CHORO << 84/06/07,
LEICESTER UNIVERSITY

of economic status and mobility describe a transient population normally found in areas with older and cheaper accommodation close to the city centre. Figure 8.3 shows high component scores in the central zone and also in isolated areas to the east, south and north-west of the study area.

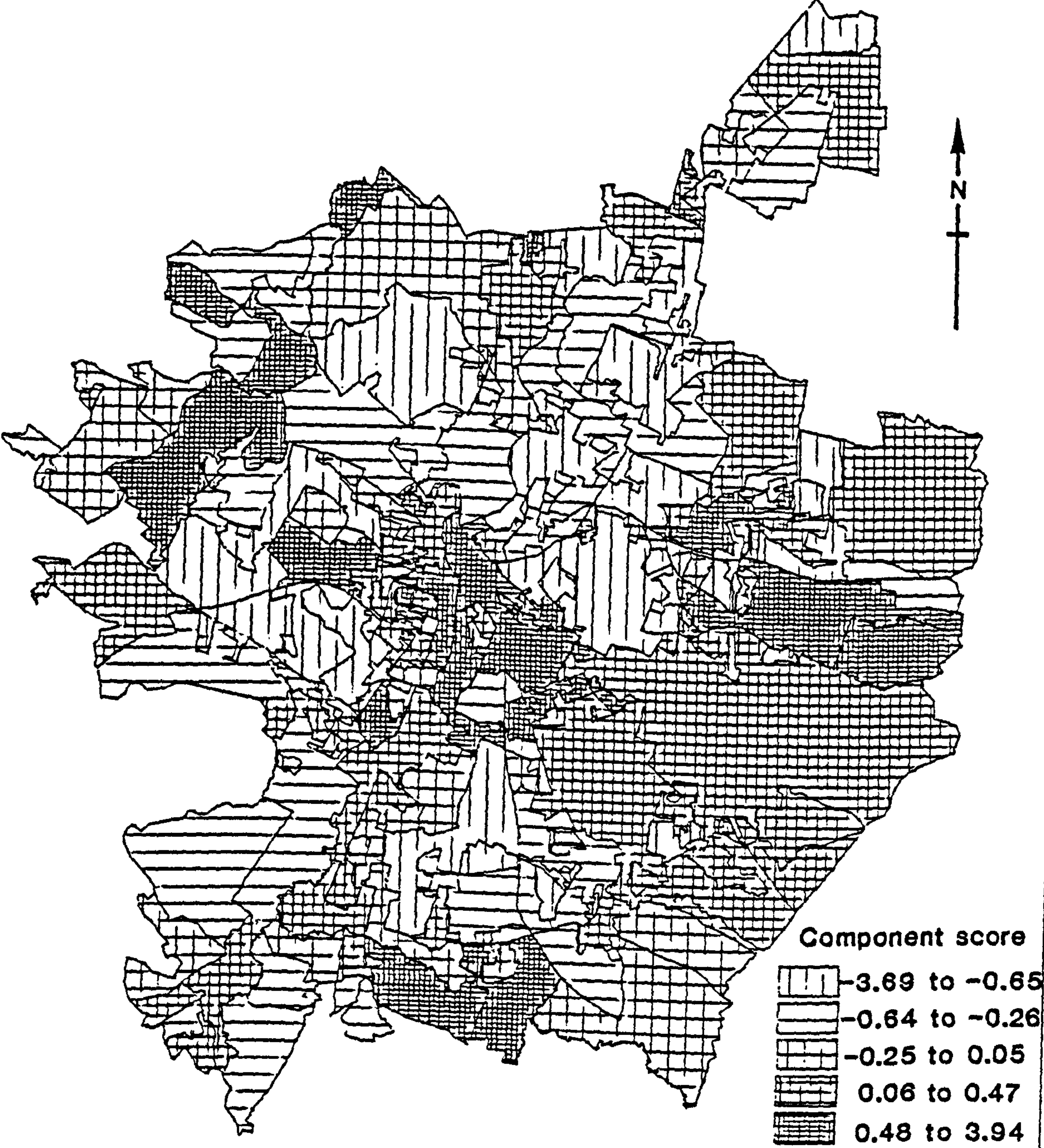
Employment indicators constitute three of the highest four loadings on component 4, with self-employed persons and students taking positive values, and the economically active population having a relatively high negative loading. Other significant variables are mixed families, which is positively loaded, and purpose-built flats, having a negative loading. In conjunction with Figure 8.4 these suggest that the component distinguishes between areas of differing socio-economic status.

Components 5 and 6 have certain common characteristics, in particular, both have the 45 to retirement age group as the highest loading variable. Both also have significant loadings from the skilled manual socio-economic group and from households lacking full amenities, however these take negative values on component 5 and positive values on component 6. Overall, component 5 is indicative of housing tenure, with a positive loading on council housing and negative loadings on owner-occupation and private renting, whilst component 6 better describes housing conditions, as shown by fairly high weightings on purpose-built flats and shared accommodation.

Only four variables have loadings greater than 0.25 on component 7. Of these, the semi-skilled manual occupational group and the immigrant population have high positive loadings, and the 5 to 15 year age group and households with many children have slightly lower negative loadings. Both these and the distribution of scores shown in Figure 8.7 suggest that the component describes patterns of ethnicity in the study area.

Figure 8.3

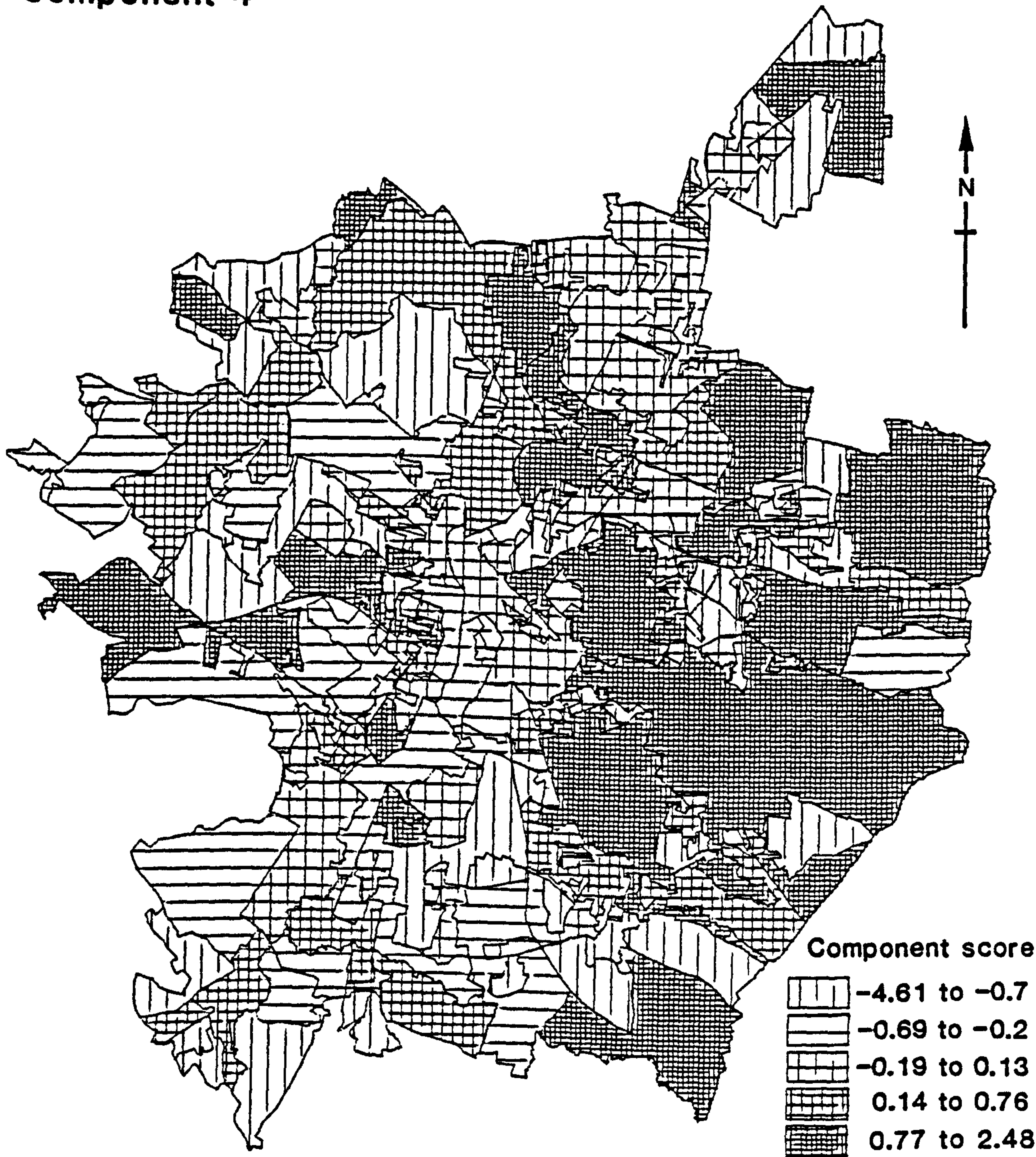
Component 3



PLOTTED BY >> CHORO << 84/06/07.
LEICESTER UNIVERSITY

Figure 8.4

Component 4

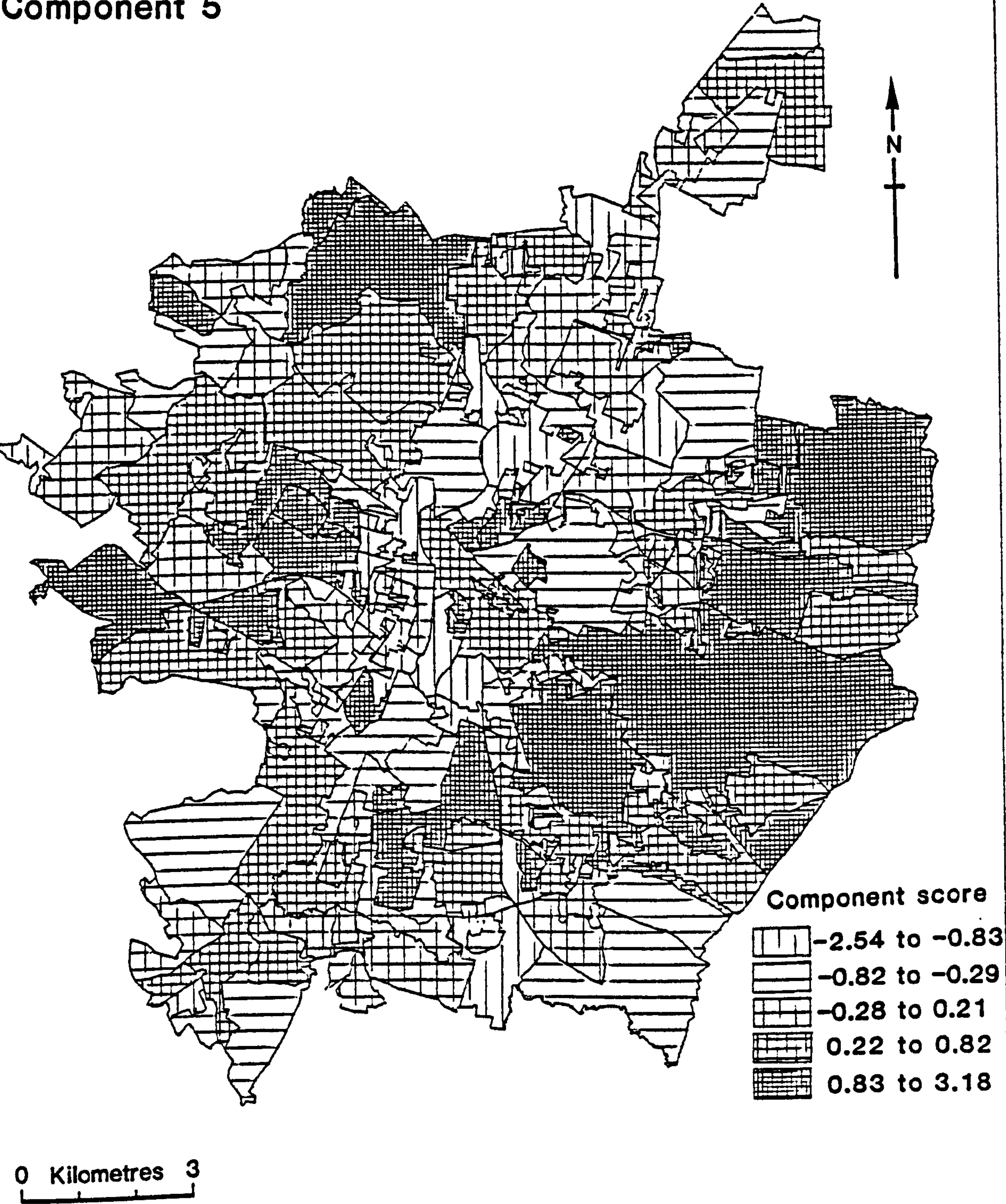


0 Kilometres 3

PLOTTED BY >> CHORO << 84/05/08.
LEICESTER UNIVERSITY

Figure 8.5

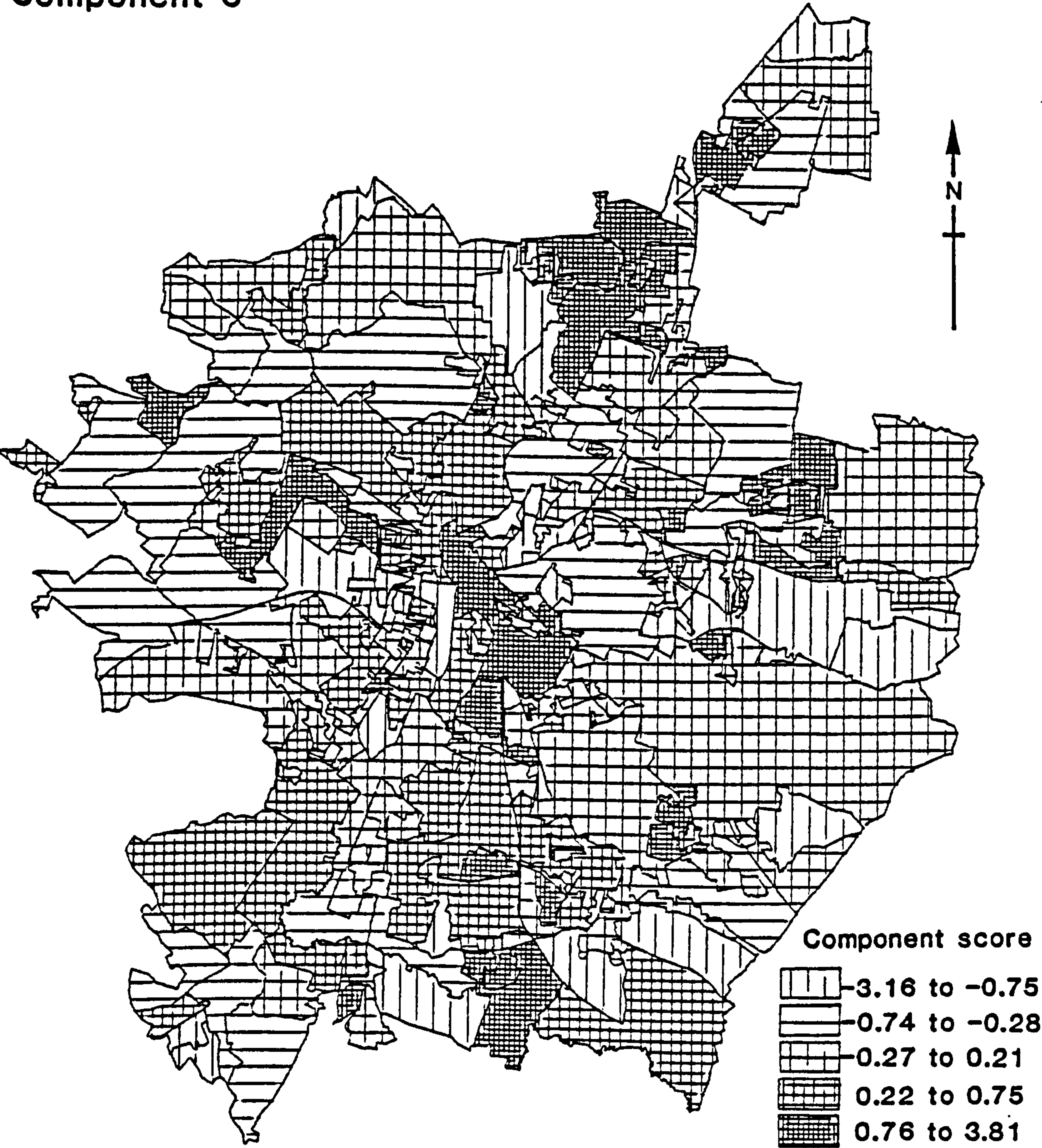
Component 5



PLOTTED BY >> CHORO << 84/06/07.
LEICESTER UNIVERSITY

Figure 8.6

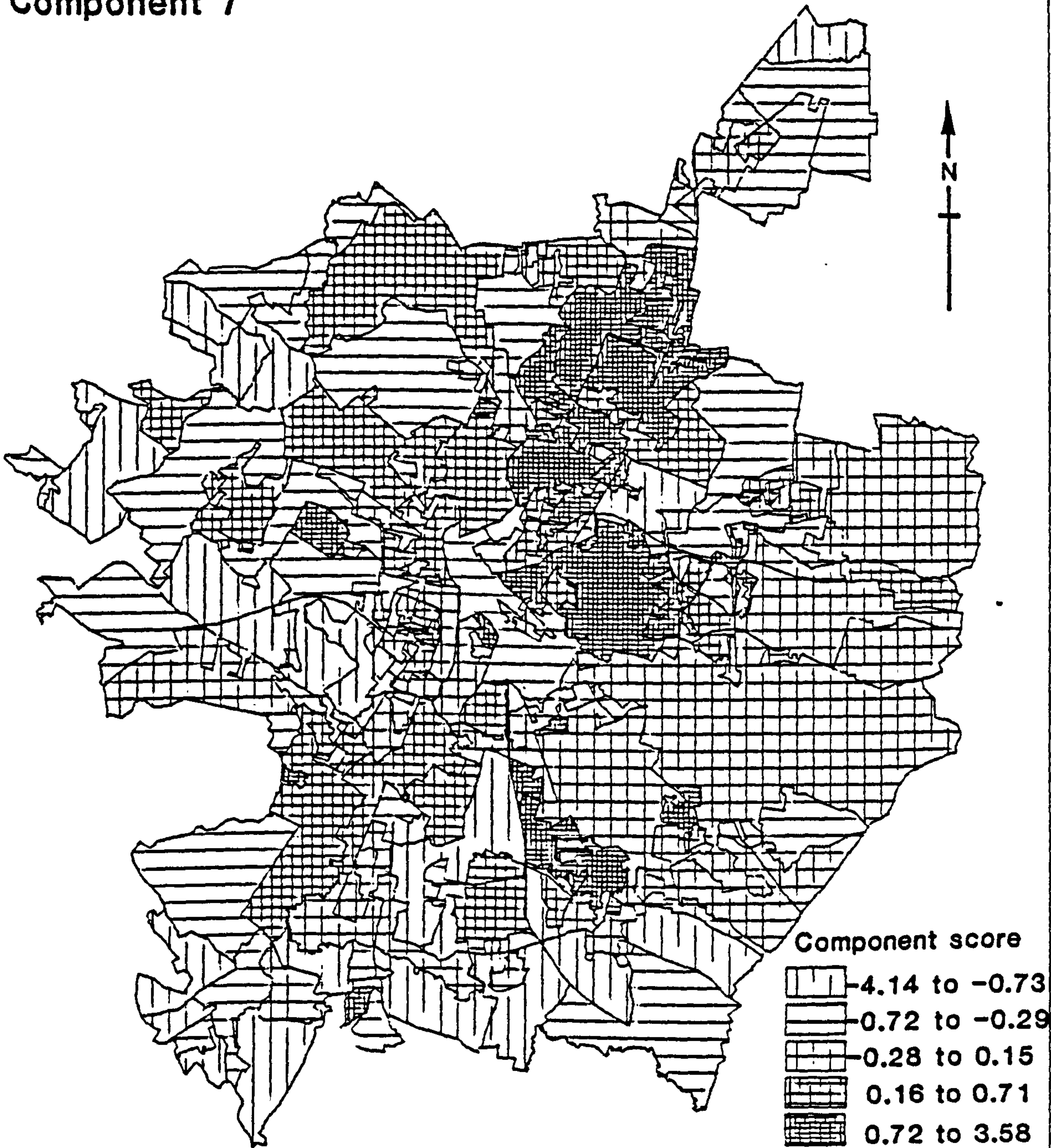
Component 6



PLOTTED BY >> CHORO << 84/06/07.
LEICESTER UNIVERSITY

Figure 8.7

Component 7



PLOTTED BY >> CHORO << S4/06/07.
LEICESTER UNIVERSITY

Several comments arise from the above descriptions. First, since the sizes of the component loadings decrease markedly as the components are extracted it becomes increasingly difficult to describe the later components. Secondly, there appears to be considerable overlap between the variables which describe each component, for example components 5 and 6 have several similar characteristics. Both the duplication and the difficulty experienced in labelling the components are unsatisfactory in a PCA and indicative of a poor result. Finally, there is an element of uncertainty about the data used which might have an unpredictable effect on the results of the PCA, this arises from the fact that the areal units used in this PCA were created by a previous cluster analysis which in turn was based on a PCA of the same set of variables.

Given these problems it is therefore not surprising that the component maps and the perinatal mortality rate (Figure 7.7) bear little relationship. For this exercise it was decided that a simple visual comparison of pairs of 'percentile' shaded maps would be appropriate, allowing as it does a direct comparison of highest and lowest ranking areas. Table 8.2 summarises the expected and observed relationships between census and perinatal data. The PCA has produced maps which are very complicated and are difficult both to interpret and explain. For example, component 3, which was labelled comparatively easily as 'transient population', exhibits no systematic association with perinatal mortality. This might be because there genuinely is no association between the two, or it might be the result of conflicting influences within the component. For example the 'positive' influence of the young and well-educated contrasts with the 'negative' influence of the economically disadvantaged, both of

Table 8.2 Comparing the distributions of component scores and perinatal mortality : expected and observed results.

Component	Main Positive loading variables	Relationship with perinatal mortality	
		Expected	Observed
1 'Work status'	Non-car ownership Single person households Council housing Pensioners	Some positive association with this general indicator of disadvantage	Some correspondence in a positive direction but with notable exceptions the outer suburbs
2 'Life cycle'	Overcrowding Young people Unemployment Ethnicity Migrancy	Positive association	No systematic association
3 'Economic status and mobility'	Private renting Shared accommodation Post school education Migrancy Single person households	Uncertain due to potentially conflicting effects of component variables	No systematic association
4 'Socio-economic status'	Self employed Students Mixed families	Negative correlation	Negative association shown in some areas
5 'Housing tenure'	45 to retirement age group Council housing Young unemployed Post school education	None, or positive association	No apparent relationship
6 'Housing conditions'	45 to retirement age group Skilled manual occupations Households lacking amenities Shared accommodation Private renting	Slight positive relationship	If any, negative
7 'Ethnicity'	Semi-skilled occupations New Commonwealth born Married women at work Purpose-built flats	Positive association	No systematic association

these groups may be found in close proximity in the inner area where scores for this component are high.

In conclusion, the PCA has failed to produce any further explanation of the pattern of perinatal mortality in Greater Leicester. The advantages that the technique normally offers by way of data reduction and simplification have not been realised in this application, not least because the input data are perhaps inappropriate to this type of processing.

8.2.2 Correlation coefficient for areal association

A less sophisticated statistic for comparing the two data sets is the correlation coefficient for areal association. This is calculated from the product moment correlation coefficient formula given below, where a and b are areal data values from the two data sets.

$$r = \frac{1/n \left(\sum (a - \bar{a})(b - \bar{b}) \right)}{\sigma_a \cdot \sigma_b}$$

Table 8.3 shows the results of three calculations, using different data values to represent census and perinatal data. Column 1 shows coefficients based on the simplest absolute values i.e. a = the number of perinatal survey cases in each area; b = the number of observations with the characteristic in each area (e.g. the number of owner-occupied households); and mean values are the sum of either of the above divided by the number of areas (i.e. 227). Since both a and b here are both related to a third variable - area size - a strong positive correlation always results, which has little to do with the relationship being tested. Column 2 avoids this problem by standardising the data through the use of ratio values i.e. a = number of mortalities/number of births, by area; b = number of observations with the characteristic/number of observations, by area (e.g. number of owner-occupied households/all households); and mean values are the sum of the

Table 8.3 Correlation coefficients for areal association - comparing the distribution of perinatal mortality with census variables

Census variable	Type of Measurement		
	Absolute	Ratio	Chi-square
DYNGCH	.925	- .166	.136
DOLDCH	.847	- .083	.138
HHSING	.762	.149	.100
HHCHIL	.948	.004	.089
HHLPAR	.654	.066	.148
HHXFAM	.918	- .005	.096
HSOWN	.632	- .114	.102
HSCOUN	.331	.031	.103
HSRENT	.814	.160	.093
HSAMEN	.888	.113	.091
HSCRWD	.945	.073	.074
HSSHAR	.466	.120	.095
HSFLAT	.209	- .025	.104
EMEACT	.789	.009	.124
EM2ACT	.766	- .095	.101
EMUNEM	.958	.114	.111
EMSTU	.749	- .009	.123
EMMWOM	.697	.030	.103
EMUNYG	.942	.041	.206
EMSEMP	.640	- .030	.120
ETNCWP	.924	.082	.110
MMIG	.906	.090	.137
MNOCAR	.894	.135	.098
SEPROF	.300	- .037	.108
SENMAN	.500	- .042	.104
SESMAN	.815	.001	.101
SESS	.892	- .104	.099
SEUS	.848	.122	.097
SEEDUC	.331	.070	.150
DYNGWK	.887	.089	.137
DMIDWK	.827	.011	.138
DOLDWK	.684	.039	.140
DYNPNS	.689	.020	.140
DOLPNS	.682	.071	.139

ratios divided by the number of areas. Although a range of both positive and negative correlation coefficients is now produced they fail to reach any level of significance. Column 3 shows correlation coefficients based on signed chi-square values for census variables and mortalities. These were previously found to be more appropriate than ratios for mapping purposes, but appear to be inappropriate in this calculation - resulting almost exclusively in low positive correlations. The chi-square value is based on the ratio value and deviance from the ratio mean and as a result suffers from similar problems to that statistic. Thus, the number of areas having few or zero mortalities skews any distribution of raw or transformed data and renders the use of a parametric test such as this inappropriate.

8.2.3 Risk surface

A third attempt at combining the two data sets was the calculation and plotting of a 'risk surface', that is, a map in which contours join points of equal risk of adverse perinatal outcome. Two procedures were followed. The first involved the calculation of a surface based on mortality data alone. This was seen as an 'observed' surface, describing the actual pattern of mortalities over the survey period, but not necessarily having any predictive value. The second approach was to take 'risk factors', i.e. variables from either the census or the perinatal survey shown to be associated with adverse perinatal outcome, and create a theoretical surface based on the distribution of these. The main advantage of the latter approach was that it could be calculated independantly of the actual mortality pattern and therefore in principle be applied to other areas.

In practice the calculation of both mortality and risk surfaces was unsuccessful, again largely due to the irregular

distribution of the perinatal mortality data. The use of the GHOST80 graphics package for plotting a contoured surface required that the data be input for regular grid squares, thus mortalities and subsequently births and risk variables were each allocated to one kilometre squares. This resulted in a large number of squares with zero data values, having either no mortalities, no births, or simply no data. This in turn caused problems for the plotting of the surface, particularly at the boundaries of the study area (see Figure 8.8). Taking mortalities as an example, a maximum of 65 cases were allocated to a single grid square and a minimum of 0. For some squares a zero value represented the true number of mortalities, for the majority however, the square lay outside the limits of the study area and no data were present. Without further processing GHOST80 could not distinguish between the two types of zero and therefore could not make allowance for 'missing' data when plotting the surface.

A second problem encountered when plotting the mortality surface was the wide variation in mortality values between adjacent grid squares. This resulted overall in a sharp, uneven surface with isolated peaks and troughs reflecting extreme values and the resulting map took on an ambiguous geometric style (see Figure 8.9).

In order to calculate a surface based on 'risk variables' it was decided that a 'risk index' should be created. Several approaches to this were taken. The simplest was to identify a two-way classification of perinatal survey variables - 'risk' or 'no risk' - based on the findings of Chapter 7. Each case or control was given a potential risk score according to the number of risk categories it fell into (see Table 8.4). These scores were then accumulated by grid square to give a value for plotting, the end result being a risk map for population as represented by

Table 8.4 Calculating a risk index from Perinatal data - risk categories

Variable	Risk categories
NAGE	< = 16, > = 35
MHITE	< = 150
PAR1	0, > = 4
GRAVID	1,> = 7
OUTC1 - OUTC7	<50, >50
MCLASS	4, 5, 8
MSEG	10, 11
PSEG	7, 10, 11, 15
METHG	3, 7
MOOB	7, 8, 9, 10, 11
MRELIG	1
IWAIT	< = 1000, 1001 - 1500, 1501 - 2000
LPREG	< = 36
LABOUR	3, 4
PRESNT	2
TYPECC	2, 4, 5, 6
NATT2	0, 1-5
NIPANC	> 20
CLINIC	0
OBSTET	1

Figure 8.8

Perinatal mortality rate

Surface plot by grid square

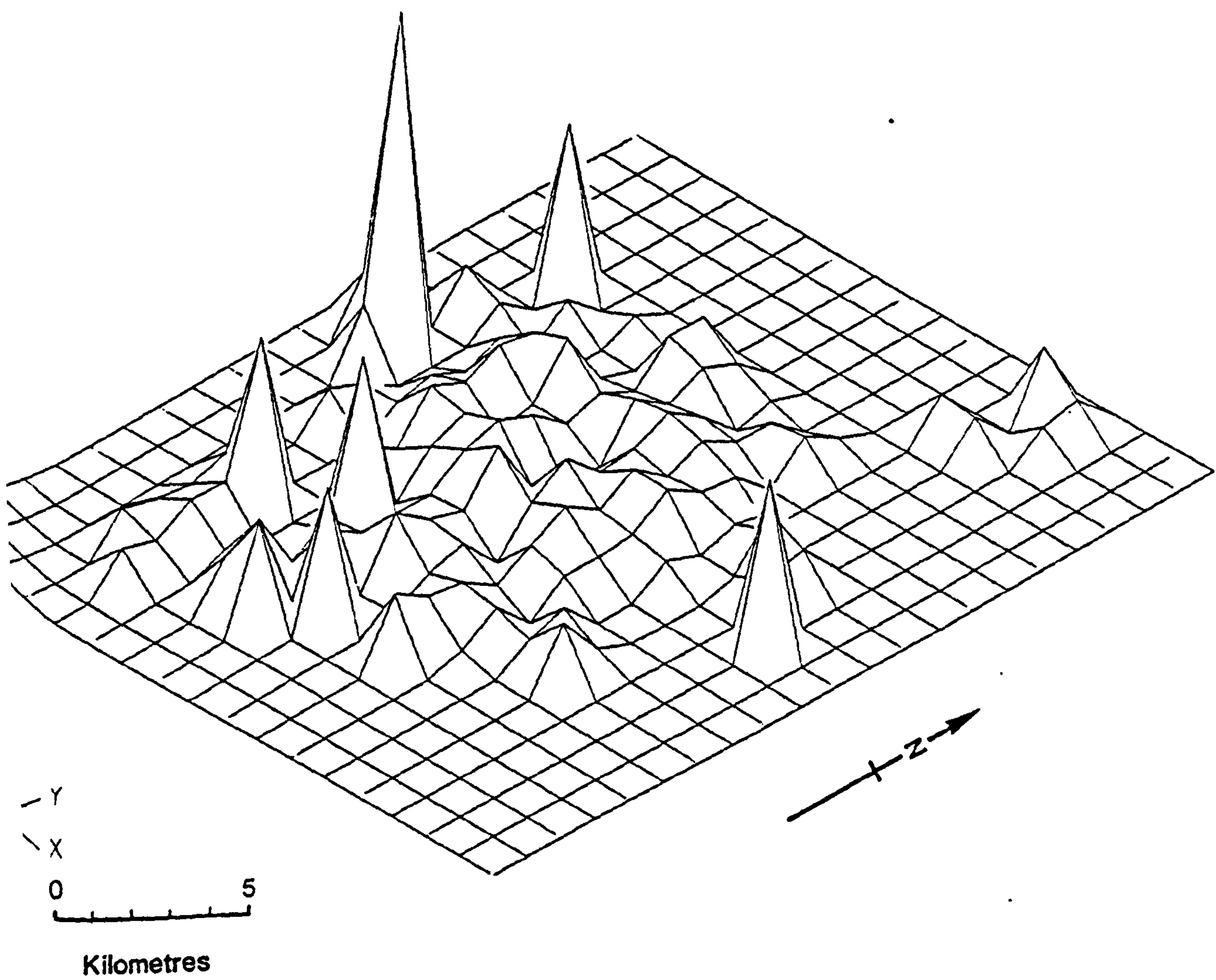
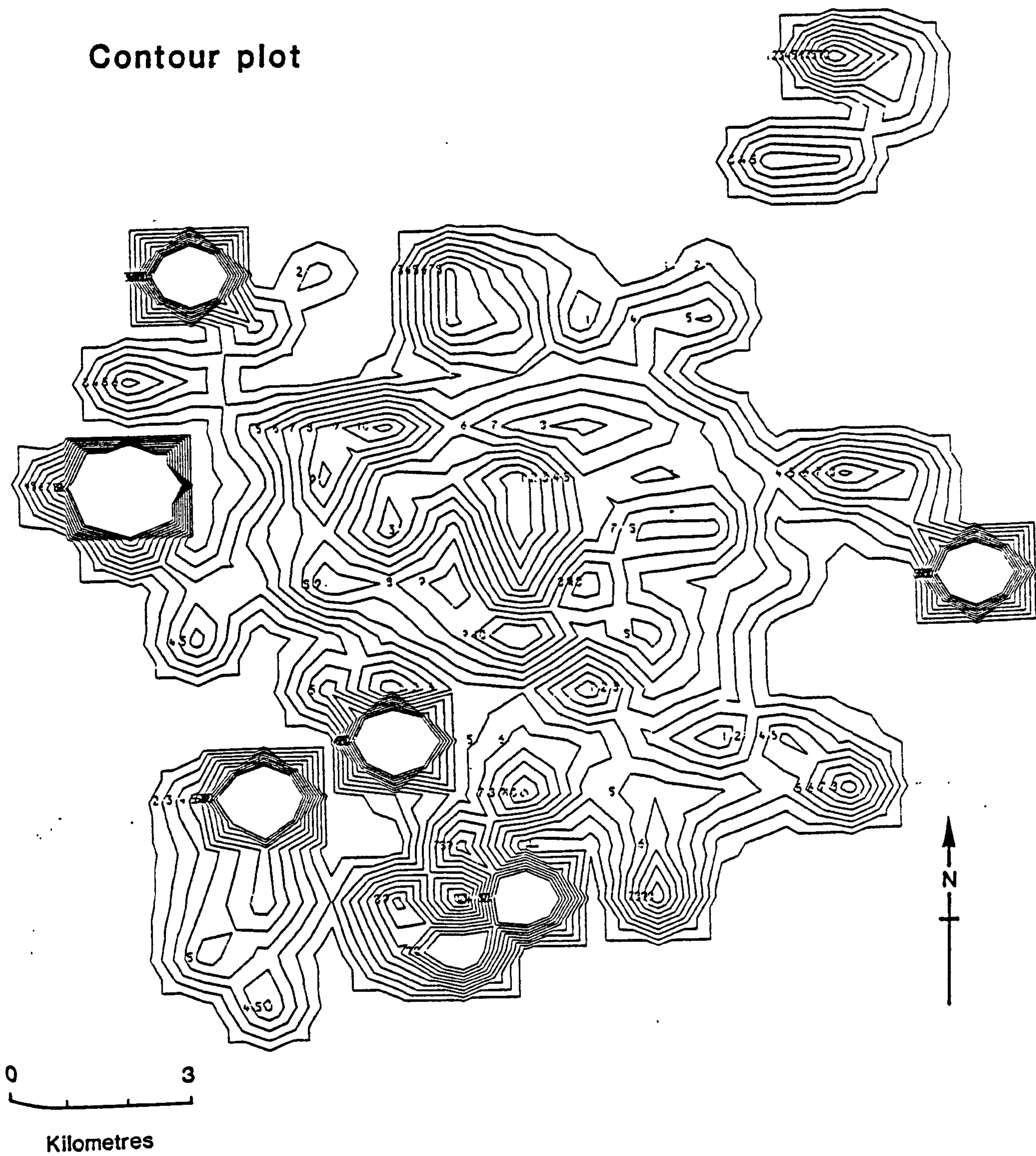


Figure 8.9

Perinatal mortality rate

Contour plot



the perinatal survey (see Figures 8.10 and 8.11).

A second 'risk index' was also based on the survey data. The same variables were used, but a 'weighting' scheme was employed such that cases or controls falling into the very high risk categories (for example, birthweight less than 1000g, no attendances for antenatal care, previous perinatal loss and so forth) should be given higher risk scores. Whilst more accurately reflecting the risk situation, this caused additional problems in deciding the relative weightings of the different classes.

The third attempt at calculating a risk index employed data from the census. Using a similar method to the above, risk scores were calculated for census areas based on the chi-square classification of selected variables. Two points are noteworthy. First, census variables describe the social and economic environment into which the baby is born. Whilst some categories might be associated with perinatal mortality (such as the different socio-economic groups), most would have greater impact after the mother leaves hospital, particularly in the first few months of life. For this reason the variables chosen for calculating the risk index describe areas in which a relatively high proportion of infants live in disadvantaged conditions, they are listed in Table 8.5. The second significant point is that this index is based ultimately on data for EDs, it is therefore a risk map for areas and in this way differs from the population risk map produced from the perinatal survey data.

Contour style plots for each of the risk surfaces were produced with the objective of comparing these with the mortality surface. Although benefitting from a greater number of observations and hence more data, the previously noted problems around peaks, troughs and the study area boundary were again

Figure 8.10

Perinatal risk index

Surface plot

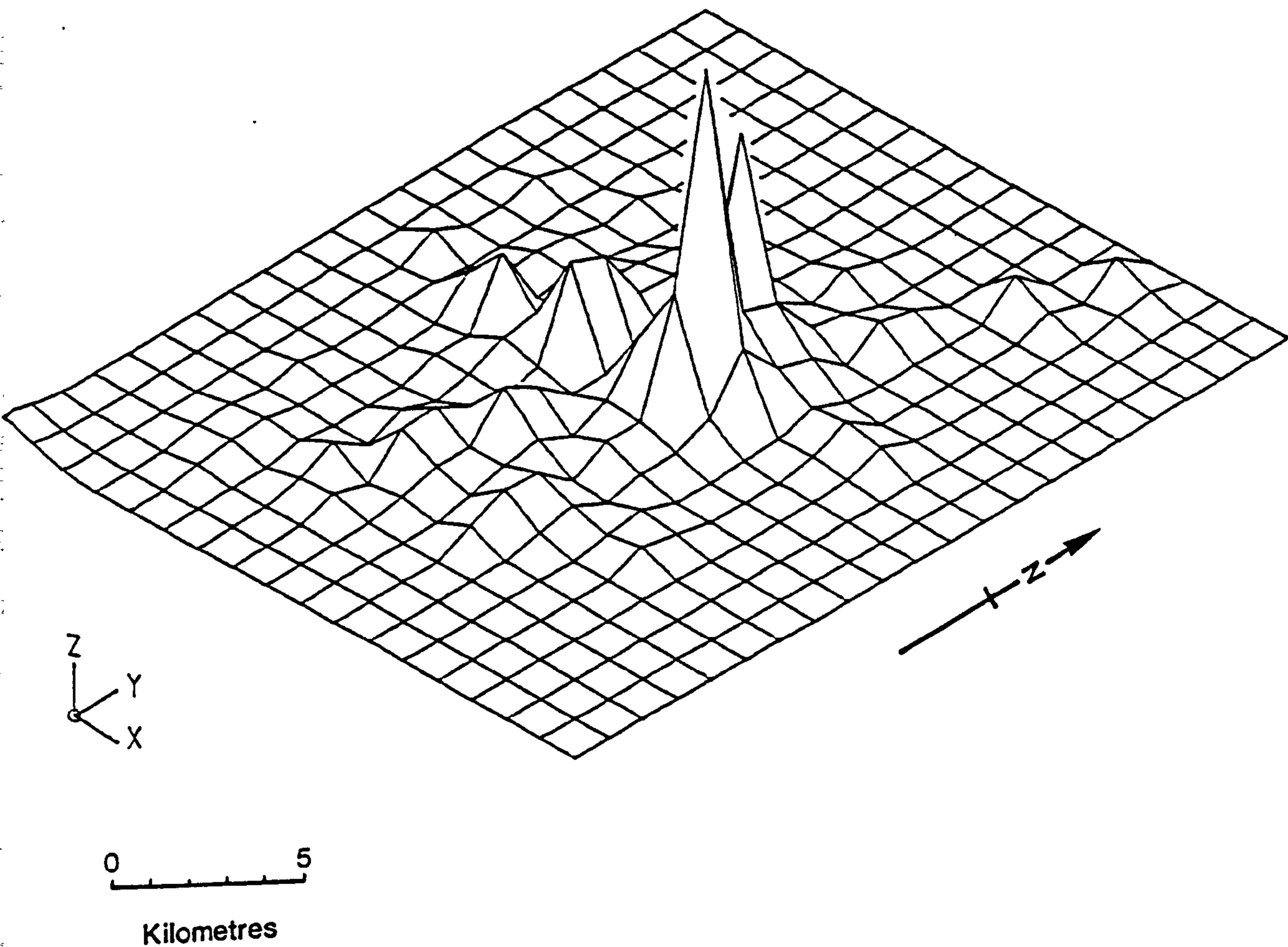
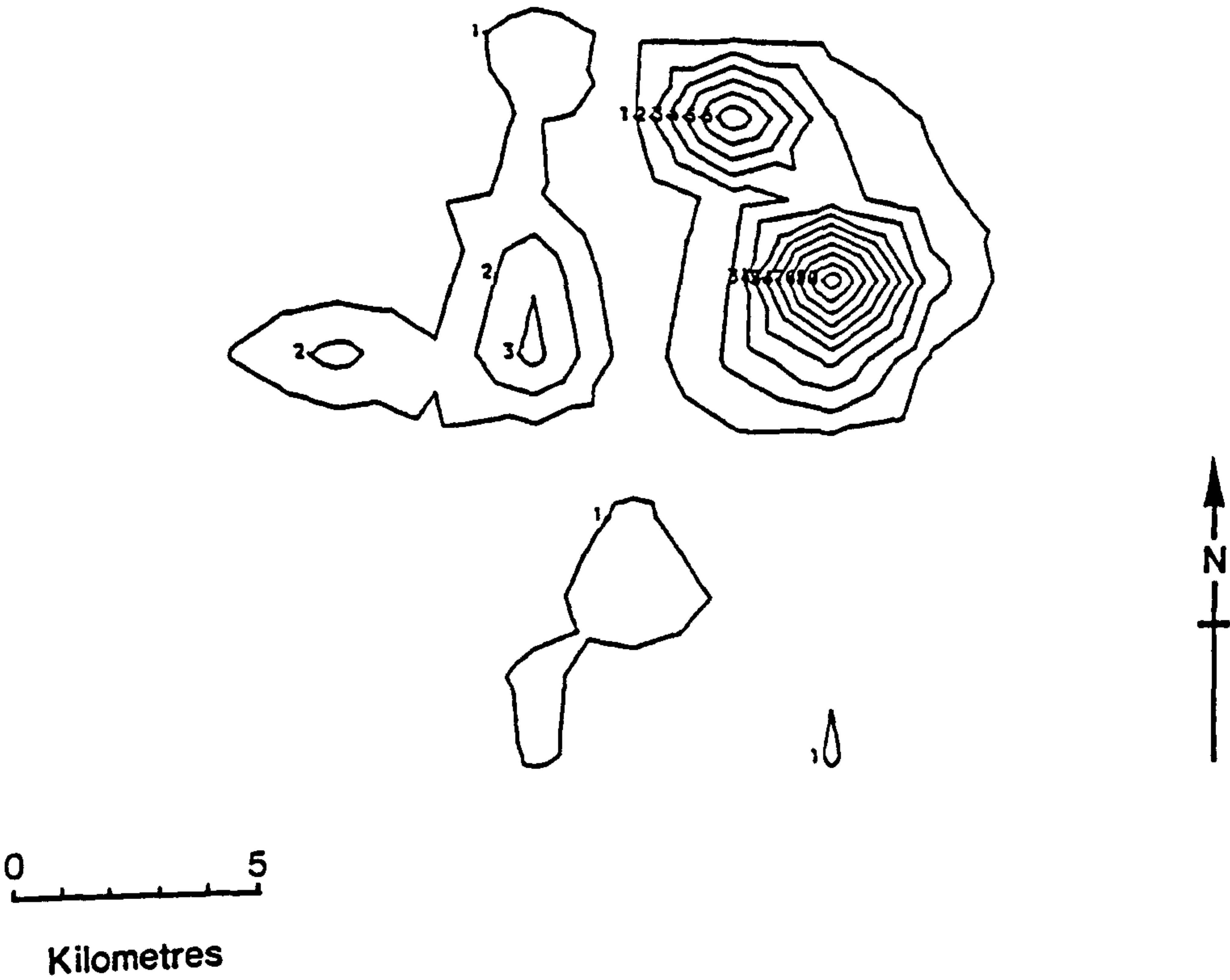


Figure 8.11

Perinatal risk index

Contour plot



encountered (see Figures 8.11 and 8.12). It was therefore concluded that the use of a risk surface, whilst interesting and potentially valuable, could at this stage provide no new information so the exercise was abandoned.

8.2.4 Log-linear methods

The attempt at calculating a risk index did however, highlight the need for an objective approach to deciding the relative importance of the different variables. Thus far, it had only been possible to compare and assess variables individually with perinatal outcome, no method of comparing these jointly to assess their relative or combined effect had been used. The following paragraphs examine the use of the log-linear model with a view to further analysing the perinatal data.

Log-linear methods are normally used for modelling the relationships along variables in multiway contingency tables, allowing both ordered and unordered categories which may vary from one variable to another. Compared with conventional multivariate techniques they make few demands on the data. The object of a log-linear model is to identify a structure which underlies the set of variables. This is achieved by applying some model to the observations such that the parameters in the model represent the 'effects' that particular variables, or combinations of variables, have in determining the values taken by the observations (Everitt, 1977). A preferred model will minimise the number of parameters required to correctly summarise the data.

Two types of model are commonly used - 'saturated' and 'non-saturated', both take the basic form,

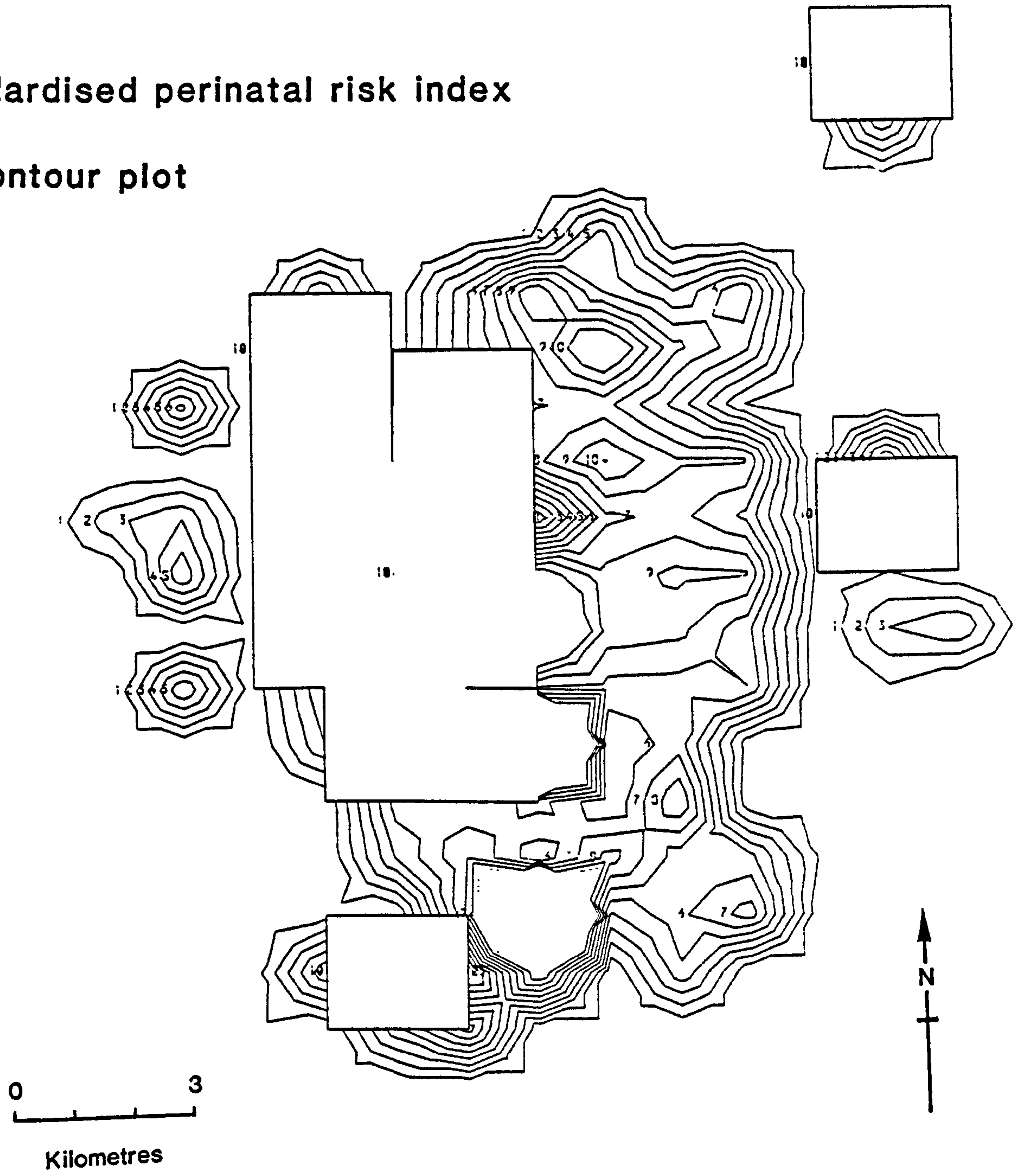
$$(i) \quad F_{ij} = \eta \tau_i^A \tau_j^B \tau_{ij}^{AB}$$

where, for a saturated model:

Figure 8.12

Standardised perinatal risk index

Contour plot



- F_{ij} is the expected frequency of cases in cell i,j which will be found if the model is true (this is a measure of the probability of an observation falling into cell i,j)
- η is the geometric mean of the number of cases in each cell in the table, used in this context as a 'baseline' from which effects are measured
- τ are the 'effect parameters' which represent the effects which variables have on the cell frequencies, they are measured as a departure from 1.0

A and B are the two variables

i and j are categories of A and B.

If any of the tau (τ) parameters are set to 1.00 the assumption is made that a particular variable has no effect on the cell frequencies, this defines the simpler, **non-saturated** model. There are four possible non-saturated models for the simplest two-variable contingency table. These are as follows:

- (ii) $F_{ij} = \eta \tau_i^A \tau_j^B$ (ie. A and B are assumed unrelated)
- (iii) $F_{ij} = \eta \tau_i^A$ (ie. categories of B are equally probable)
- (iv) $F_{ij} = \eta \tau_j^B$ (ie. categories of A are equally probable)
- (v) $F_{ij} = \eta$ (ie. all categories are equally probable)

Models (i) to (v) comprise the entire set necessary to test a variety of hypotheses concerning the two variables.

Two points arise from these equations. First, it is noted that no model contains a higher order tau (such as τ^{AB}) without also containing its lower order components (τ^A and τ^B). Put another way, if the association between two variables A and B is set to zero, then so are all the higher order associations involving A,B and one or more other variables (Upton,1981). Log-linear models are usually 'hierarchical' in this way because of constraints imposed by the maximum likelihood procedures used to calculate values for the tau parameters (see Dening and Stephan, 1940; Goodman, 1970 and 1971; and Everitt, 1977 for further details of these calculations).

The second point concerns the 'fit' of the model to the data. In the saturated model the number of parameters exactly

equals the number of cells, therefore the model fits the data perfectly. The expected values are simply the observed frequencies. The non-saturated model, however, has had certain parameters removed. It is therefore likely to produce expected frequencies which are more or less divergent from the data (Knoke and Burke, 1980). The goodness of fit of the model may be tested using either the chi-square statistic,

$$\chi^2 = \sum (f_{ij} - F_{ij})^2 / F_{ij}$$

or more commonly, the likelihood-ratio statistic,

$$L^2 = 2 \sum f_{ij} \ln (f_{ij} / F_{ij})$$

where

f_{ij} = observed cell frequencies

F_{ij} = expected cell frequencies

(Upton and Fingleton (1979) note that in practice the results of these two statistics will usually be very similar). To find the best fit model the test statistic must be minimised - the larger it is relative to the degrees of freedom (d.f. = the number of tau parameters set to 1.00), the greater the difference between observed and expected values and therefore the less adequate the representation of the relationships among the variables.

Thus far, all the models shown have been multiplicative. By taking natural logarithms of all the terms, the equations can be converted into linear equations, hence the term 'log-linear' model. For example, equation (i) would take the form

$$\begin{aligned} \ln (F_{ij}) &= \ln (\eta \tau_i^A \tau_j^B \tau_{ij}^{AB}) \\ &= \ln (\eta) + \ln (\tau_i^A) + \ln (\tau_j^B) + \ln (\tau_{ij}^{AB}) \end{aligned}$$

which is otherwise written as

$$(vi) \quad G_{ij} = \theta + \lambda_i^A + \lambda_j^B + \lambda_{ij}^{AB}$$

where $\theta = \ln (\eta)$, $\lambda = \ln (\tau)$ and $G_{ij} = \ln (F_{ij})$.

Similar expressions can be created for non-saturated models and for models of tables with multiple dimensions. Unfortunately in

descriptive terms the multiway table produces cumbersome log-linear models which incorporate large numbers of parameters. For example the saturated model for a three-way table may be written as follows

$$(vii) G_{ijk} = \theta + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_{ij}^{AB} + \lambda_{jk}^{BC} + \lambda_{ik}^{AC} + \lambda_{ijk}^{ABC}$$

As the number of variables and/or categories increases, so too do the number of cells in the table and the likelihood of zero cell frequencies. Eventually it is possible for the number of cells to exceed the number of observations, guaranteeing zero values. Zero frequencies arising in this way (they are known as 'sampling zeros' to distinguish them from 'fixed zeros' which occur when a particular combination of characteristics does not occur in the population) cause two problems (Upton, 1978). First, the assumptions of the likelihood-ratio statistic become severely strained and secondly, it becomes impossible to fit the saturated model (the logarithm of 0 being $-\infty$).

A second consideration is that as the number of dimensions increases, so too does the number of possible non-saturated models. Upton (1981) suggests that it is helpful to examine the parameters of the saturated model first to see which may be excluded, if this still leaves several possible models (as is likely) these may be tested using one of the test statistics.

The general log-linear model makes no distinction between independent and dependent variables, for this the more specialised logit model is required. The logit model uses variables that are classified as either 'explanatory' (i.e. independent) or 'response' (i.e. dependent); the objective of applying such a model is to examine the effects of individual, and combinations of, explanatory variables on the response variables. No account is taken of the interactions between responses.

Logit models, and their application in geography are fully described by Wrigley (1976,1985). Examples of equivalent logit and log-linear models are given by Haberman (1978).

For the analysis of the perinatal data the 'PLR' program for stepwise logistic regression from the BMDP statistical package was chosen. This program distinguishes between dependant and independant variables, selecting the latter in a stepwise manner and estimating the coefficients as described above. At each step, one variable is entered or removed from the model depending on the value of either a maximum likelihood ratio (Wrigley, 1985 describes how this may be calculated) or an approximate asymptotic covariance estimate. Although the former more frequently appears in the literature, the latter is considerably more economical in computing time (being approximately eight times as fast) and was found to give very similar results, it was therefore used as the basis for the models described below. A simple hierarchical rule was initially implemented - by default only one term could be moved at one step. It could be entered if all its lower order interactions were already in the model, or removed if none of its higher order interactions were present.

At each step, the log-likelihood, change in log-likelihood and three goodness-of-fit statistics are provided. The first is the likelihood-ratio statistic as described above, the second is the 'Hosmer' goodness-of-fit test which compares observed and expected frequencies of ten cells (see Hosmer and Lemeshow, 1981), and the third is the 'C.C. Brown' goodness-of-fit test which compares the fit of data to the logistic or to some alternate member of the family of models as defined by Prentice (1976). Small probability values for the latter tests indicate a poor fit to the data and an inappropriate choice of the logistic model respectively. (Engelman, 1981).

Initially, six 'models' were specified for analysis. These differed with respect to the independent variables included in each model. Model 1 included all of the variables previously found to be of significance when tested using the chi-square statistic, and the remaining models included different selections of these variables. In each model an effort was made to include variables likely to have some common influence on perinatal outcome. For example, model 2 included variables such as maternal age and height, parity, gravidity, social class, socio-economic group, country of birth, ethnic group, religion, cluster membership and previous pregnancy outcome. Values for each of these were determined before the start of the pregnancy and were viewed as fixed for the duration. Model 4, in contrast, contained only measures of health care availability and uptake, both of which could be influenced by the actions of health care providers during the pregnancy. In the specifications of models 3 and 5 there was a high degree of overlap, both included obstetric details pertaining to the present pregnancy (for example, gestational age, fetal presentation and type of labour), however model 5 included some additional historical information (such as parity and previous pregnancy outcome). The final model included only variables of social and geographical interest; cluster membership, social class, the socio-economic group of both parents, country of birth, ethnicity and religion comprise the full set. The purpose of this model was to assess how accurately these variables alone could predict perinatal outcome.

No interactions were nominated for any of the models, the objective being the assessment of the relative importance of the different factors. In each case perinatal outcome, coded as dead or alive, was the dependent variable. The variables used are

summarised in Table 8.6, the results are shown in Tables 8.7 to 8.12, and the critical probabilities at the final step are shown in Table 8.13.

Examining firstly the critical final probabilities, only in models 3 and 4 are there sufficient observations relative to data cells to fully justify the use of the likelihood ratio statistic (shown in the tables under the more general label of 'goodness-of-fit chi-square'). Nevertheless, the following comments can be made. First, the very low C.C. Brown probabilities for models 2 (characteristics of the mother) and 6 (socio-geographic characteristics) suggest that the logistic model is inappropriate to these subsets of the data, a disappointing result from a geographical point of view since in both cases cluster membership was the first variable to be entered into the model. Second, marginal C.C. Brown probabilities and low Hosmer probabilities for both model 3 (characteristics of the pregnancy and infant) and model 5 (obstetric history) indicate a poor fit of predicted to actual values, a result which is supported in model 3 by a low probability for the likelihood ratio value. Third, although the C.C. Brown probability suggests the logistic model is probably appropriate to the subset of health care variables in model 4, the two other goodness-of-fit measures differ substantially, neither however indicates a particularly good fit. The most useful of the six models is therefore model 1 which includes all the variables previously found to be of significance. All three goodness-of-fit measures have high final probabilities indicating that for this combination of variables, the logistic model is both appropriate and accurate at predicting perinatal outcome. The results of this model will be examined further.

Eight variables were entered into model 1, none were removed. In order of entry, and hence relative importance, these

Table 8.6 Stepwise Logistic Regression : Variables included
in the analysis

Variable Name	Variables included :					
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
NAGE	✓	✓				
PAR1	✓	✓			✓	
CLUSTI	✓	✓				✓
TYPECC	✓		✓		✓	
GRAVID	✓	✓			✓	
NATT2	✓			✓		
NIPANC	✓			✓		
MHITE	✓	✓				
LPREG	✓		✓		✓	
PRESNT	✓		✓		✓	
LABOUR	✓		✓		✓	
OBSTET	✓			✓		
IWAIT	✓		✓		✓	
CLINIC	✓			✓		
MCLASS	✓	✓				✓
MSEG	✓	✓				✓
PSEG	✓	✓				✓
MCOB	✓	✓				✓
METHG	✓	✓				✓
MRELIG	✓	✓				✓
OUTCM	✓	✓			✓	
OUTCA	✓	✓			✓	
No. of cases	244	251	671	669	670	255
No. of controls	<u>207</u>	<u>207</u>	<u>684</u>	<u>680</u>	<u>684</u>	<u>208</u>
Total responses	451	458	1355	1349	1354	463
No. of cells	447	441	213	104	651	271

Step Variable Log-likelihood

Improvement

Goodness-of-fit

Hosmer

C.C. Brown

	entered	Likelihood		Chi-square		Chi-square		goodness-of-fit		goodness-of-fit	
		DF	Prob	DF	Prob	DF	Prob	DF	Prob	DF	Prob
0		-311.090				622.18	446	0		0	1
1	LPREG	-231.153		2	0	153.873	444	.265		0	1
2	IWAIT	-188.933		6	0	84.44	438	.983	1.92	7	.424
3	LABOUR	-178.432		7	.004	21.003	431	.996	4.635	8	.581
4	GRAVID	-169.761		9	.044	17.342	425	.999	8.038	8	.774
5	PSEG	-162.335		5	.011	14.852	420	1.0	2.861	8	.735
6	OUTCM	-160.155		1	.037	4.36	419	1.0	3.021	7	.601
7	NIPANC	-153.107		6	.029	14.097	413	1.0	1.618	7	.661
8	CLINIC	-150.764		1	.030	4.685	412	1.0	1.579	7	.702

Parameters at the final step:

Term	Coefficient	Standard Error		Term	Coefficient	Standard Error	
GRAVID(1)	-3.3599	1.230	*	LPREG (1)	-0.7873	0.273	*
(2)	-2.0965	1.246		(2)	0.2390	0.430	
(3)	-2.1035	1.317		LABOUR(1)	-0.0856	0.317	
(4)	-2.7123	1.405		(2)	0.6127	0.405	
(5)	-----	-----		(3)	1.4958	0.578	*
(6)	-----	-----		(4)	0.3766	1.038	
(7)	-4.6291	5.867		(5)	-1.7753	0.422	*
(8)	2.575	9.146		(6)	0.4248	0.494	
(9)	-----	-----		(7)	-0.2285	0.443	
NIPANC(1)	-0.4109	0.483		PSEG (1)	1.4380	0.431	*
(2)	0.1802	0.651		(2)	-0.1006	0.286	
(3)	0.7755	0.585		(3)	0.3979	0.322	
(4)	-1.7918	0.998		(4)	-0.2897	0.645	
(5)	-2.2909	0.954	*	(5)	-1.6317	0.503	*
(6)	3.2472	1.397	*	IWAIT (1)	0.3795	0.479	
CLINIC	-0.4233	0.198	*	(2)	-0.0244	0.365	
OUTCM	0.8690	0.429	*	(3)	-0.5873	0.302	
CONSTANT	2.5953	1.278	*	(4)	-1.1765	0.349	*
				(5)	-1.3440	0.432	*
				(6)	-0.9540	0.653	

*Terms for which coefficient/standard error exceeds 2.0

Table 8.8

Model 2 Characteristics of the mother - all determined before she starts the pregnancy

Step	Variable entered	Log-likelihood	Improvement Chi-square	DF	Prob	Goodness-of-fit Chi-square	DF	Prob	Hosmer goodness-of-fit	DF	Prob	C.C. Brown goodness-of-fit	DF	Prob
0		-315.344				611.280	440	0				0	0	1
1	CLUSTI	-302.226	26.236	8	.001	585.044	432	0	0	6	1	0	0	1
2	MSEG	-292.700	19.053	5	.002	565.992	427	0	4.481	8	.811	3.324	2	.190
3	MRELIG	-289.315	6.769	3	.08	559.222	424	0	5.386	8	.716	3.624	2	.163

Parameters at the final step:

Term	Coefficient	Standard Error	Term	Coefficient	Standard Error
CLUSTI(1)	0.1421	0.296	MSEG (1)	-0.6968	0.302 *
(2)	0.9911	0.651	(2)	0.6064	0.363
(3)	0.4678	0.236	(3)	-0.1827	0.295
(4)	0.4691	0.326	(4)	0.6799	0.520
(5)	-2.3260	0.976 *	(5)	-0.4378	0.259
(6)	-0.9456	0.426 *	MRELIG(1)	-0.4363	0.220
(7)	0.4207	0.523	(2)	0.0768	0.226
(8)	1.0722	0.400 *	(3)	-0.0021	0.196
			CONSTANT	0.1017	0.273

Table 8.9

Model 3 Characteristics of the pregnancy and of the infant

Step	Variable entered	Log-likelihood	Improvement		Goodness-of-fit Chi-square		Hosmer Goodness-of-fit		C.C. Brown Goodness-of-fit	
			Chi-square	DF	Prob	Chi-square	DF	Prob	Chi-square	DF
0		-939.152				1053.905	212	0		
1	LPREG	-656.876	564.548	2	0	489.351	210	0		
2	IWAIT	-538.103	237.546	6	0	251.808	204	.013	1	1
3	LABOUR	-518.569	39.067	7	0	212.740	197	.210	.830	.830
4	PRESNT	-511.838	13.464	2	.001	199.276	195	.402	.416	.783
									.336	1.470
										.479

Parameters at the final step:

Term	Coefficient	Standard Error	Term	Coefficient	Standard Error
LPREG (1)	-0.4236	0.136 *	PRESNT (1)	0.3299	0.292
(2)	-0.4310	0.215 *	(2)	0.3955	0.419
LABOUR(1)	-0.0843	0.177	IWAIT (1)	0.6863	0.284 *
(2)	-0.0661	0.194	(2)	0.5937	0.216 *
(3)	1.3087	0.316 *	(3)	-0.4066	0.158 *
(4)	-0.1667	0.524	(4)	-1.2936	0.176 *
(5)	-0.7437	0.193 *	(5)	-1.0953	0.198 *
(6)	0.2068	0.247	(6)	-1.3394	0.338 *
(7)	0.1501	0.227	CONSTANT	0.7983	0.259 *

Table 8.10

Model 4 Health care variables – level and quality of health care provision

Step	Variable	Log-likelihood	Improvement		Goodness-of-fit		Hosmer		C.C. Brown	
			Chi-square	DF	Prob	Chi-square	DF	Prob	Goodness-of-fit	DF
0		-935.011				523.128	103	0	0	0
1	NATT2	-735.212	399.597	3	0	123.531	100	.055	0	0
2	CLINIC	-731.141	8.143	1	.004	115.388	99	.125	2.618	6
3	NIPANC	-724.768	12.746	6	.047	102.642	93	.232	5.610	8

Parameters at the final step:

Term	Coefficient	Standard Error
NATT2 (1)	-0.2211	0.135
(2)	-1.2990	0.151 *
(3)	-0.4919	0.324
NIPANC(1)	-0.1575	0.178
(2)	0.3762	0.211
(3)	0.3947	0.262
(4)	-0.3406	0.349
(5)	0.0231	0.337
(6)	0.0225	0.326
CLINIC	-0.2608	0.098 *
CONSTANT	0.3072	0.167

Table 8.11

Model 5 Mother's obstetric history - including both past and present conditions

Step	Variable	Log-likelihood	Improvement		Goodness-of-fit		Hosmer		C.C. Brown	
			Chi-square	DF	Chi-square	DF	goodness-of-fit	DF	goodness-of-fit	DF
0		-938.452			1380.319	650	0		0	1
1	LPREG	-656.803	563.298	2	817.014	648	0	1	0	1
2	IWAIT	-538.085	237.435	6	579.586	642	.963	3.591	.368	2
3	LABOUR	-518.550	39.070	7	540.516	635	.997	8.180	.488	2
4	PRESNT	-511.829	13.443	2	527.073	633	.999	9.073	1.462	2

Parameters at the final step:

Term	Coefficient	Standard Error	Term	Coefficient	Standard Error
LPREG (1)	-0.4235	0.136 *	PRESNT(1)	0.3293	0.292
(2)	-0.4305	0.214 *	(2)	0.3958	0.419
LABOUR(1)	-0.0844	0.177	IWAIT (1)	0.6864	0.284 *
(2)	-0.0661	0.194	(2)	0.5939	0.216 *
(3)	1.3086	0.316 *	(3)	-0.4064	0.158 *
(4)	-0.1667	0.524	(4)	-1.2930	0.175 *
(5)	-0.7438	0.193 *	(5)	-1.0952	0.198 *
(6)	0.2068	0.247	(6)	-1.3394	0.338 *
(7)	0.1500	0.227	CONSTANT	0.7980	0.259 *

Table 8.12

Model 6 Socio-geographic characteristics

Step	Variable	Log-likelihood	Improvement Chi-square	Goodness-of-fit Chi-square	Hosmer goodness-of-fit	C.C. Brown goodness-of-fit
			DF Prob	DF Prob	DF Prob	DF Prob
0		-318.537		402.043 270 0		0 0 1
1	CLUSTI	-305.158	8 .001	375.285 262 0	0 6 1	0 0 1
2	MSEG	-296.059	5 .003	357.087 257 0	5.899 8 .659	3.467 2 .177

Parameters at the final step:

Term	Coefficient	Standard Error	Term	Coefficient	Standard Error
CLUSTI(1)	0.0523	0.291	MSEG (1)	-0.6653	0.293 *
(2)	1.0070	0.647	(2)	0.6860	0.356
(3)	0.4600	0.222 *	(3)	-0.1145	0.284
(4)	0.3731	0.317	(4)	0.5999	0.511
(5)	-2.3239	0.977 *	(5)	0.5412	0.249 *
(6)	-0.9141	0.426 *	CONSTANT	0.2355	0.256
(7)	0.3708	0.519			
(8)	1.1296	0.395 *			

Table 8.13

Models 1 to 6 : critical probabilities at the final step

Model	Final Step	Cases Cells	Goodness-of-fit chi-square	Hosmer goodness-of-fit	C.C.Brown goodness-of-fit
1	8	451/447	1.0	0.979	0.702
2	3	458/441	0.0	0.716	0.163
3	4	1355/213	0.402	0.336	0.479
4	3	1349/104	0.232	0.691	0.859
5	4	1354/651	0.999	0.336	0.481
6	2	463/271	0.0	0.659	0.177

were

- (i) length of the pregnancy (gestational age)
- (ii) infant birthweight
- (iii) type of labour
- (iv) gravid state
- (v) father's socio-economic group
- (vi) previous perinatal death
- (vii) number of days in-patient antenatal care
- (viii) availability of an antenatal clinic session at the G.P.'s surgery

A total of 12 categories of these variables were found to make a significant contribution to the model, these each had coefficients at least twice as large as their standard errors. The coefficients may be interpreted as measures of relative risk, where each coefficient (n) in Tables 8.7 to 8.12 describes the relative probability of a perinatal death for a member of variable category (n + 1) compared with a member of variable category (1). Thus a negative coefficient indicates that the probability of a perinatal death in the higher category is smaller, a positive coefficient indicates that the risk is higher. For model 1 the following statements may therefore be made.

- The probability of an infant born after 36 to 41 weeks gestation dying is less than that for an infant born before 36 weeks gestation.
- The probability of an infant weighing either 3000g to 3500g or 3500g to 4000g dying is less than that for an infant weighing less than 2000g.
- The probability of a mother that has had medical induction (prostaglandins) experiencing a perinatal death is greater than that for a mother who did not labour, but the probability for a mother that has had augmented labour (ARM) is less.
- The probability of a mother in her second pregnancy experiencing a perinatal death is less than that of a mother in her first pregnancy.
- The probability of a perinatal death is greater for infants with fathers in socio-economic groups 5 and 6 (intermediate and junior non-manual) when compared with those with fathers in groups 1 to 4 (employers, managers and professionals) but less for infants with fathers in groups 12 to 15, 17 and 77 (own account, farmers and agricultural workers and students).
- A mother who has experienced a previous perinatal mortality is more likely to lose her current baby than one who has not.

- The probability of a mother receiving 20 to 30 days in-patient antenatal care experiencing a perinatal death is less than that of a mother receiving less than two days inpatient care, but the probability for a mother receiving over 30 days is greater.
- When an antenatal clinic session is available at the mother's G.P.'s surgery the probability of a perinatal death is lower than when there is not.

Most of these results support the findings of Chapter 7, the only exception is the variable 'number of days in-patient antenatal care' for which the unexpected result is that 20-30 days in-patient care carries an apparently lower risk than less than 2 days care. This can be explained by the very small number of responses in the former category (13 in total).

Having established the relative importance of the individual variables, an attempt was made to assess their combined effect on perinatal outcome. A series of models were specified as before with both full sets and subsets of variables, and interactions were specified between groups of variables. It was neither appropriate nor practical (in terms of computing time) to assess the effect of every possible combination of variables, however those which were chosen were felt to be representative of the sets most likely to make a significant contribution to perinatal outcome. Examples included interactions among the 'obstetric' variables (see models 7,8,10 and 11 in Table 8.14) and those among the 'socio-geographic' variables (models 9 and 12 in Table 8.14). Although most of these models gave high values for the goodness of fit statistics, few combinations of variables had significant coefficients at the final step. It was felt that no new information could be derived from these.

A final attempt was made to suppress the overwhelming effect of the major obstetric variables by treating the length of pregnancy as the dependant variable (with 36 weeks gestation as the cutpoint) and excluding infant birthweight from the analysis.

Table 8.14 Stepwise Logistic Regression : Assessing combinations of explanatory variables

Model: 7
Variables: NAGE, PAR1, CLUSTI, TYPECC, GRAVID, NATT2, NIPANC, WHITE, LPREG, PRESNT, LABOUR, OBSTET, IWAIT, CLINIC, MCLASS, MSEG, PSEG, MOOB, METHG, MRELIG, OUTCM, OUTCA

Interactions specified: TYPECC * LPREG * PRESNT * LABOUR * IWAIT

Final step: 0

Responses: 451

Cells: 7

Goodness-of-fit - chi-square: 0.999

Hosmer goodness-of-fit: 1.0

Brown goodness-of-fit: 1.0

Significant parameters at the final step: None

Model 8
Variables: As Model 7

Interactions specified: PAR1 * GRAVID * OUTCM * OUTCA,
TYPECC * LPREG * PRESNT * LABOUR * IWAIT

Final step: 0

Responses: 451

Cells: 30

Goodness-of-fit - chi-square: 0.9

Hosmer goodness-of-fit: 0.716

Brown goodness-of-fit: 0.839

Significant parameters at the final step: None

Model 9
Variables: As Model 7

Interactions specified: CLUSTI * MCLASS * MSEG * PSEG * MOOB
* METHG * MRELIG

Final step: 0

Responses: 451

Cells: 45

Goodness-of-fit - chi-square: 1.0

Hosmer goodness-of-fit: 1.0

Brown goodness-of-fit: 0.995

Significant parameters at the final step: CLUSTI * MCLASS * MSEG * PSEG,
PSEG * METHG * MRELIG

Table 8.14 (cont.)

Model 10

Variables: TYPECC, LPREG, PRESNT, LABOUR, IWAIT

Interactions specified: TYPECC * LPREG * PRESNT * LABOUR * IWAIT

Final step: 0

Responses: 1355

Cells: 7

Goodness-of-fit - chi-square: 0.998

Hosmer goodness-of-fit: 0.999

Brown goodness-of-fit: 1.0

Significant parameters at the final step: TYPECC * LPREG * IWAIT

Model 11

Variables: PAR1, TYPECC, GRAVID, LPREG, PRESNT, LABOUR, IWAIT, OUTCM, OUTCA

Interactions specified: PAR1 * GRAVID * OUTCM * OUTCA,
TYPECC * LPREG * PRESNT * LABOUR * IWAIT

Final step: 0

Responses: 1354

Cells: 33

Goodness-of-fit - chi-square: 0.834

Hosmer goodness-of-fit: 0.793

Brown goodness-of-fit: 0.566

Significant parameters at the final step: TYPECC * LPREG * PRESNT * IWAIT

Model 12

Variables: CLUSTI, MCLASS, MSEG, PSEG, MCOB, METHG, MRELIG

Interactions specified: CLUSTI * MCLASS * MSEG * PSEG * MCOB * METHG
* MRELIG

Final step: 0

Responses: 463

Cells: 45

Goodness-of-fit - chi-square: 1.0

Hosmer goodness-of-fit: 1.0

Brown goodness-of-fit: 0.999

Significant parameters at the final step: MCLASS * MSEG * PSEG * MCOB

All the other variables were specified as in model 1, no interactions were included (see Table 8.15). Five variables were entered into the model: attendances for antenatal care after 20 weeks gestation, type of labour, previous perinatal mortality, presentation at birth and the availability of an antenatal clinic at the G.P.'s surgery. However the goodness-of-fit statistics indicated both a poor fit to the data and an inappropriate use of the technique, therefore this model was abandoned.

8.3 Probability Mapping

The methods described in Section 8.2, whilst of interest in their own right, each have significant drawbacks when used here, due mainly to the nature of the underlying data. There is, however, one further technique which has been used successfully with similar medical-geographical data, that is the **Poisson Probability Map**.

In a study of the distribution of brain tumours in Poland, Choynowski (1959) first produced maps based on Poisson probabilities. In statistical terms, Poisson probabilities represent the theoretical prediction of the probabilities of points over small areas for a random process (Taylor, 1977), they are calculated as follows:

$$p(x) = \frac{e^{-\lambda} \lambda^x}{x!}$$

where

$p(x)$ is the probability that an area will contain x points

e is the exponential constant (2.71828)

λ is the expectation or Poisson mean (ie the average number of occurrences of the event).

The usefulness of this statistic lies in its ability to identify statistically significant events; in the context of the present study, to show patterns of mortalities which are non-

Table 8.15 Stepwise Logistic Regression : Changing the dependant variable

Model 13

Dependant variable: LPREG

Independant variables: NAGE, PAR1, CLUSTI, TYPECC, GRAVID, NATT2, NIPANC, MHITE, PRESNT, LABOUR, OBSTET, CLINIC, MCLASS, MSEG, PSEG, MOOB, METHG, MRELIG, OUTCM, OUTCA

Interactions specified: None

Final step: 5

Responses: 453

Cells: 449

Goodness-of-fit - chi-square: 1

Hosmer goodness-of-fit: 0.075

Brown goodness-of-fit: 0.126

Variables entered into model: NATT2, LABOUR, OUTCM, PRESNT, CLINIC

random under the Poisson distribution. To this end the parameters are as follows:

x is the observed number of perinatal mortalities in an NSA

λ is the expected number of mortalities in the same area calculated from the overall study area mortality rate and the size of the 'at risk' population within the NSA.

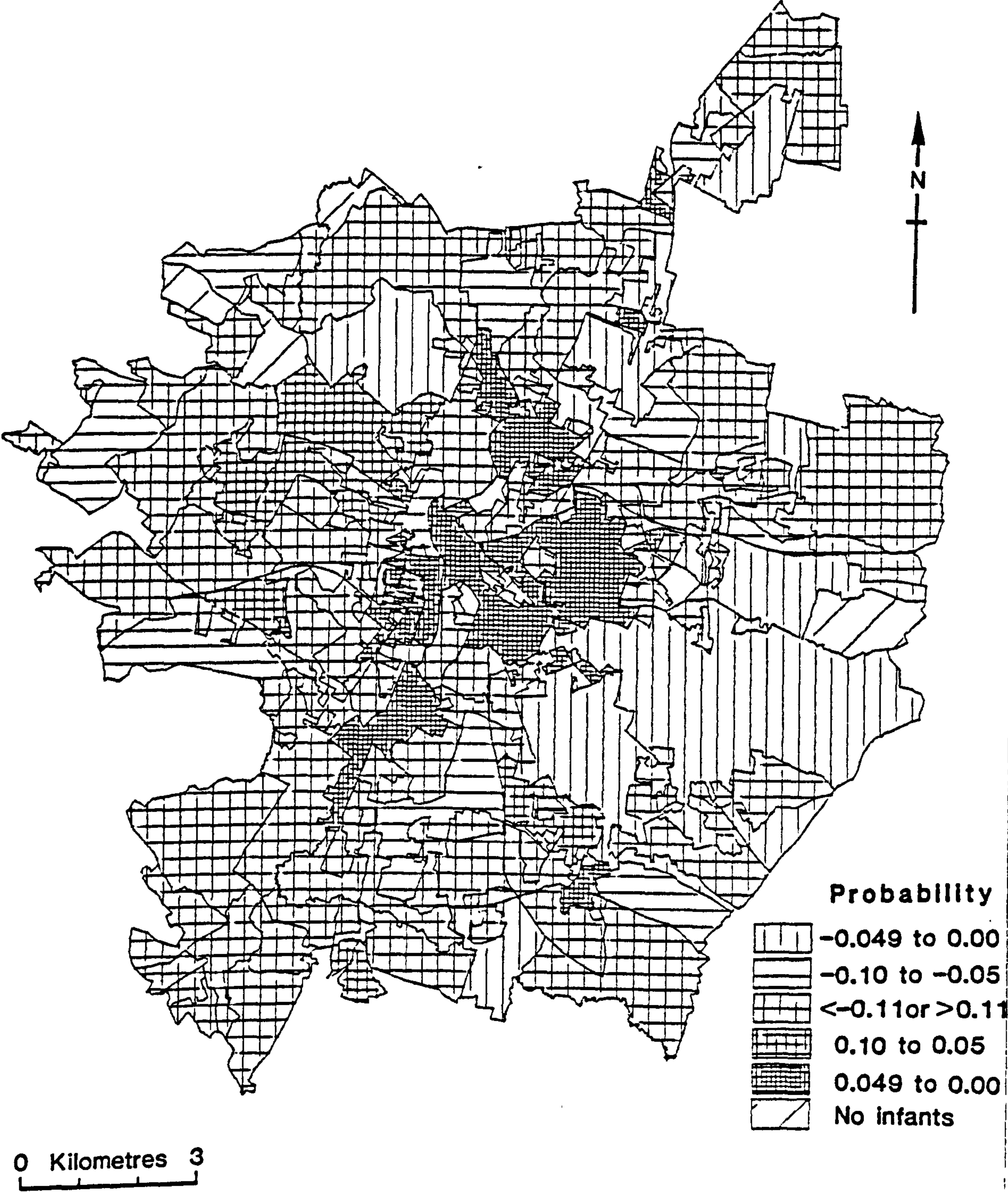
Probabilities may be either upper- or lower-tail, depending on whether the observed rate is above or below the mean rate. Significant probabilities (ie those of less than 0.1 and less than 0.05) are distinguished by appropriate shading on the maps.

The usefulness of the probability maps depends on the definition of the 'at risk' population which in turn determines the value of the expected outcome. In the present study four 'at risk' groups are identified, three from the census data namely, the infant population (aged 0-1 year), the population of young children (aged 0-4 years) and the population of married women of childbearing age (15-44 years), and one based on the Leicestershire births data set, that is the total of all births in 1980. It is noted that the definition of the 'at risk' population in this exercise assumes that it is randomly distributed. The principle of the Poisson probability map is that the underlying distribution is the result of a Poisson process with significant "outbreaks" being identified as being non-random under the Poisson distribution (Ferguson, 1977). If an attempt is made to calculate or map Poisson probabilities for mortalities against specific risk categories (for example Asian households or those headed by people in semi- and unskilled occupations) the requirement of randomness is not met and the technique is thus invalid. Figures 8.13 to 8.16 show Poisson probabilities for the four risk groups described above.

There are many similarities between the first 3 maps, all of

Figure 8.13

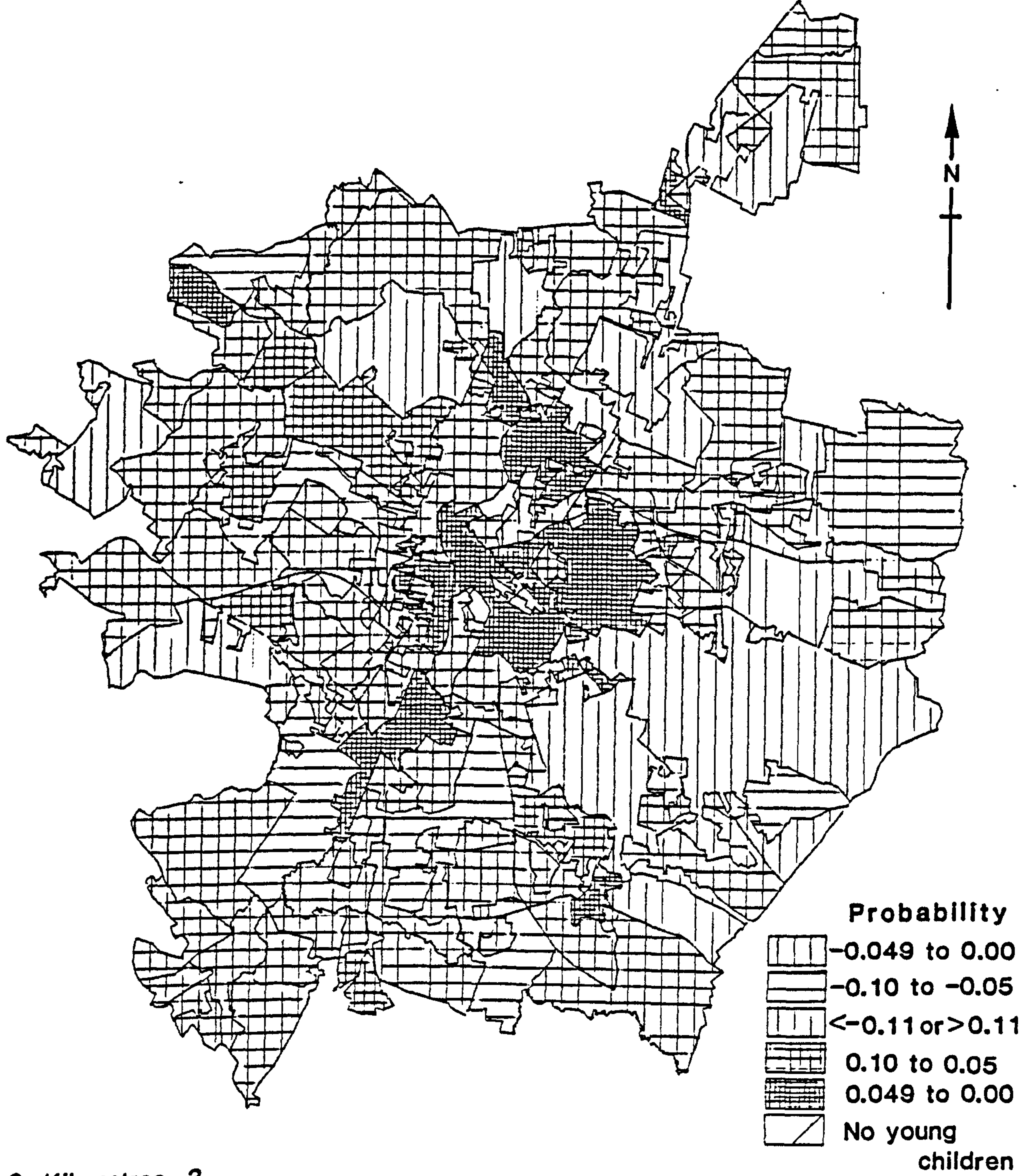
Poisson probabilities – Infant population



PLOTTED BY >> CHORO << 95/07/31.
LEICESTER UNIVERSITY

Figure 8.14

Poisson probabilities - Young children



PLOTTED BY >> CHORO << 85/07/31,
LEICESTER UNIVERSITY

Figure 8.15

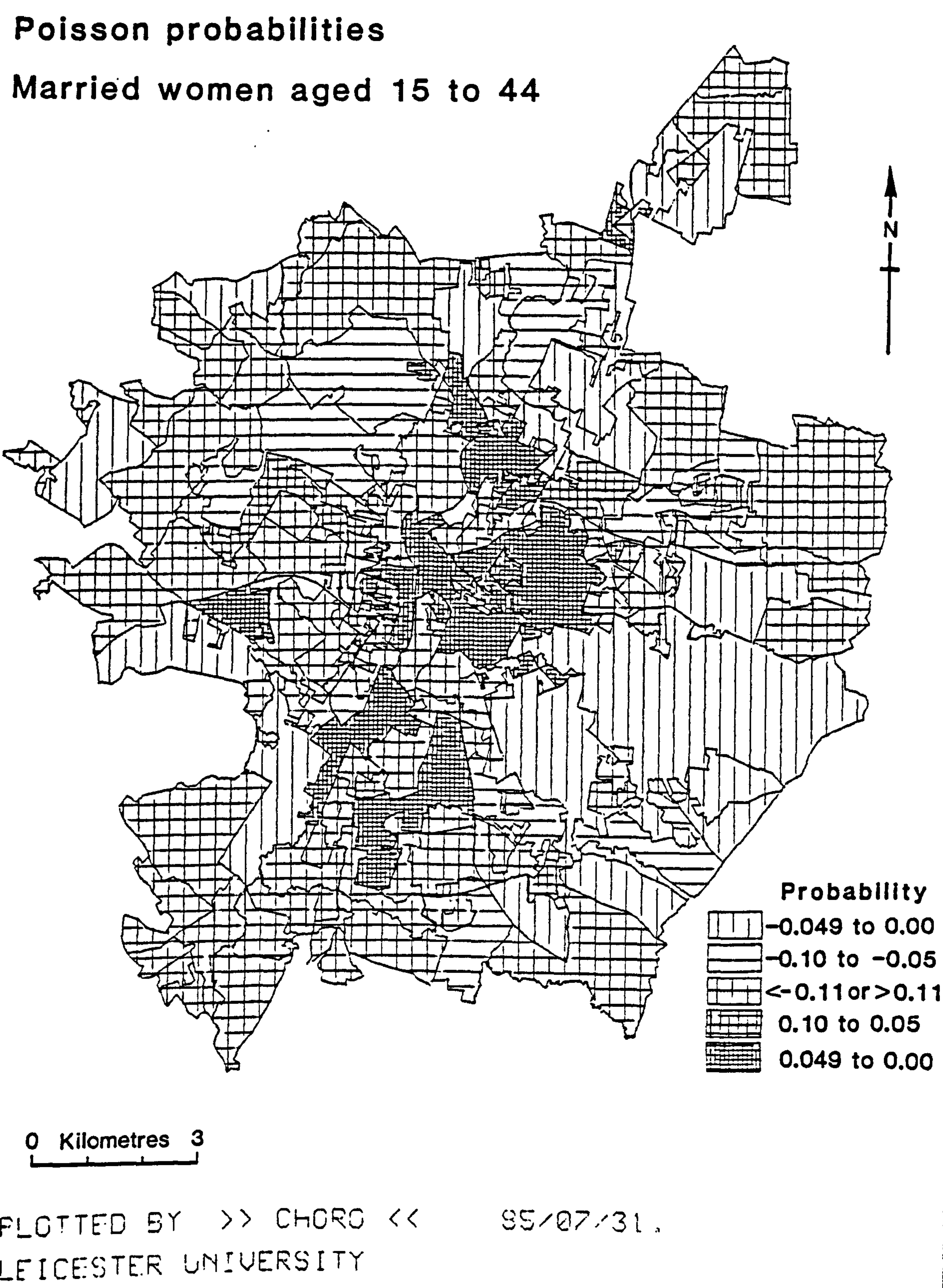
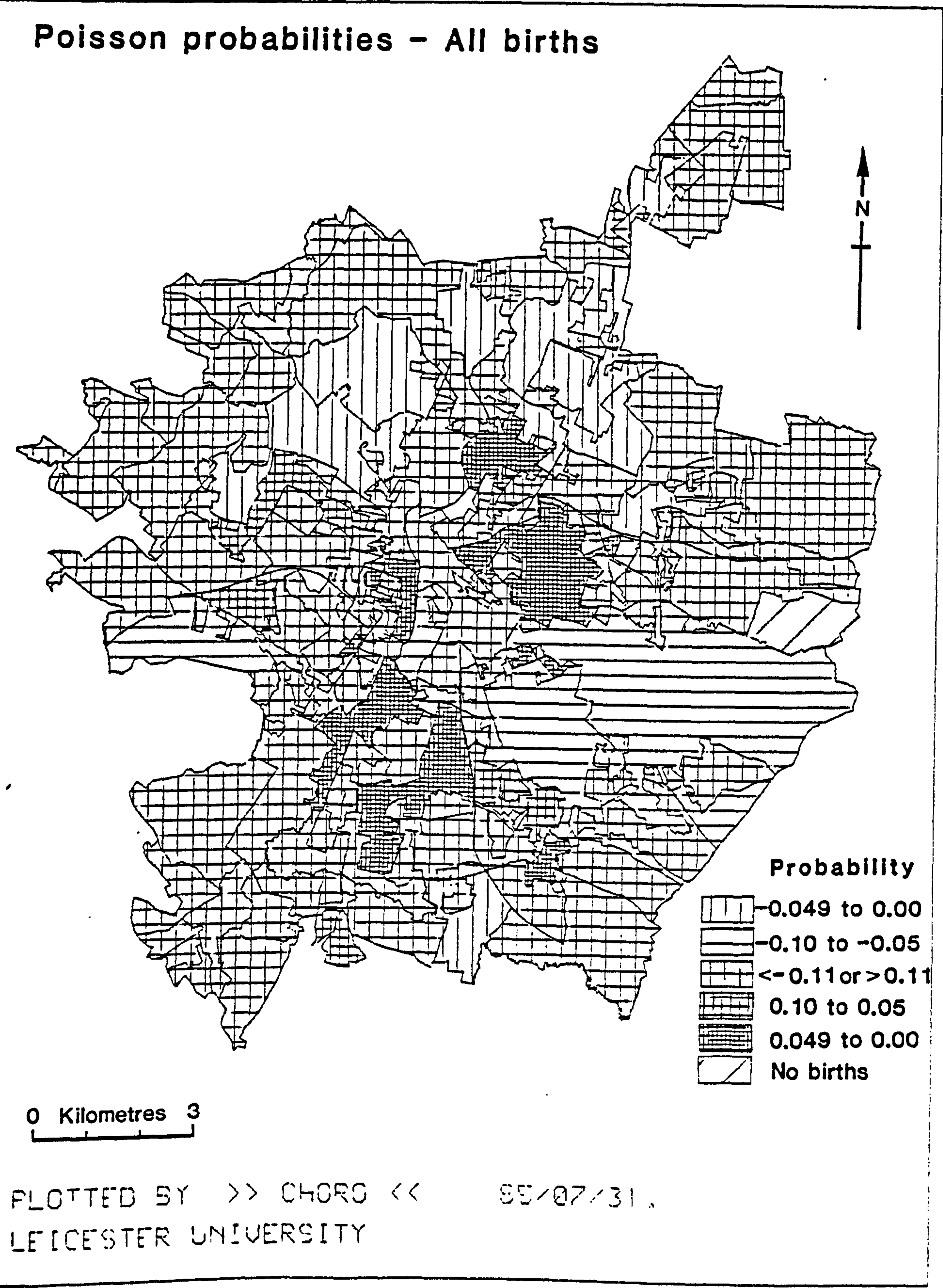


Figure 8.16



which are based on the census variables (Figures 8.13, 8.14 and 8.15). Highfields, Belgrave, the inner city and the Aylestone Road each show observed perinatal mortality rates significantly higher than would have been expected from a random distribution. In addition, the North Braunstone, Saffron Lane and Eyres Monsell pre-war council estates exhibit particularly high rates when standardised by the 'at risk' married group (Figure 8.15). On all maps the lowest rates (ie areas with probabilities in the extreme lower tail of the distribution) are found in the affluent outer suburbs, particularly to the south east of the study area.

These findings can be related to the socio-economic structure of the urban area. As previously noted (see Chapter 6) the inner city and the areas of Highfields and Belgrave are characterised by several indicators of deprivation; they have above average unemployment (Figures 6.30 and 6.33), households in shared dwellings (Figure 6.20), households lacking full amenities (Figure 6.18), lone parent households (Figure 6.13), households headed by people in unskilled or semi-skilled occupations (Figures 6.26 and 6.25) and households without a car (Figure 6.36). There are more one-year migrants (Figure 6.35) and persons born in the New Commonwealth or Pakistan (Figure 6.10) than would normally be expected. Physically, these areas are typified by high density nineteenth century housing, much of it in need of modernisation or redevelopment, congestion on the roads, a lack of open space and a generally less attractive living environment.

The other highlighted areas, namely the council estates to the west and south of the city have many similar characteristics, serving as they do society's disadvantaged groups. Lone parents, large households, the unskilled and the unemployed all have above average representation in these areas. Car ownership is low and recent population movement high.

Overall, therefore, the areas identified by the census analysis as having high levels of deprivation have also now been shown to have above average perinatal mortality rates. Figure 8.16 showing the perinatal mortality rate against the set of all births confirms these findings. Conversely, areas of higher socio-economic status such as Oadby, Stoneysgate, Stoughton and Evington to the south-east of the study area, and areas of recent expansion and development such as Birstall, Groby, Beaumont Leys, Leicester Forest East and South Wigston, have more favourable perinatal experience.

The implications of this mapping exercise are three-fold. In the first instance, perinatal mortality has been shown to be spatially concentrated not only within certain parts of the urban area but also, viewed in conjunction with the known social geography of Leicester, within specific population groups (such as the socio-economically disadvantaged) and within specific communities (for example the city's Asian community). Secondly, it has been possible to examine the relative likelihood of experiencing a perinatal death given the mother's residential location. Notwithstanding ecological fallacies this information is still invaluable in planning the spatial allocation of scarce health care resources. Finally, the association between perinatal mortality and socio-economic status, residential circumstances and the urban environment has been conclusively established. In conjunction with other known risk factors, such as those identified in Chapter 7, it should be possible now to target health care provision to both areas and people in greatest need of it.

Chapter 9. Summary and conclusions

9.1 Summary

The main objectives of this thesis have been to analyse, describe and explain the distribution of perinatal mortality within an urban area. This is a topic within medical geography on which little or no research has been undertaken, although related work has focussed on many other aspects of health in urban areas. A review of the literature points to the holistic viewpoint as being the most important principle in analysing health patterns. This approach emphasises the description and explanation of aspects of health in their widest social and environmental context. The conventional distinction between 'disease ecology' and 'the geography of health care' was felt to be an unhelpful subdivision in analysing the incidence of perinatal mortality which, as a health problem, is a product of both 'ecological' and 'health care' factors and of the relationships between them.

The discussion of perinatal mortality in Chapter 3 focuses on research undertaken by epidemiologists. In many cases strong links have been identified between this work and that by geographers, with differences being mainly of emphasis rather than subject matter or technique. Epidemiologists are better equipped both to recognise and to assess the major significance of 'medical' factors in determining perinatal outcome; geographers, on the other hand, are more directly concerned with the identification and explanation of spatial patterns of ill health. Several problems have been identified in the analysis (geographical or otherwise) of perinatal death, the most important of these being the definition of terms, the choice of survey method and type of standardisation, and the interpretation of the results, particularly the identification of causality among risk factors. Most epidemiologists simply identify 'associations'

between key variables and adverse perinatal outcome rather than offering evidence for a causal link. Factors commonly identified by epidemiologists include birthweight, gestational age, social class, maternal age, parity, legitimacy and previous adverse outcome. A range of other physical, obstetric, socio-economic and spatial characteristics are also considered important.

Assembling the main data sources used in this project, namely the 1981 Population Census, the Leicestershire Perinatal Mortality Survey and the Leicestershire Births Records, so that they could be directly compared, is one of the major practical difficulties tackled in this research. The data sets were held both on different computers and in different spatial referencing systems (EDs and postcodes), and as in each case only a subset of the original data file was required, a degree of selection both by location and of variables was needed before the data could be used. Although each data set is in itself comprehensive, indeed a census of the total population under investigation, the information available is far from ideal. The Population Census, for example, provides detailed demographic information on age and sex structure, but fails to address adequately other important issues such as fertility and ethnicity. The main deficiency in the mortality survey is that spatial elements are of secondary importance to the original design of the questionnaire, for example important factors such as accessibility to health care are not even considered. The inclusion of five additional sets of questions in the survey questionnaire, concerning the home environment, accessibility to health care, attitudes to health care, the socio-economic environment of the mother, and maternal characteristics, would have added greatly to the geographical analysis of the findings (see Table 9.1). The Leicestershire

Table 9.1 Additional questions for the Perinatal Mortality Survey

(i) The home environment

- (a) Are you an owner-occupier?
council tenant?
private tenant?
- (b) Do you live in a detached house?
semi-detached house?
terraced house?
flat?
bedsit?
- (c) Do you share your home with another family?
- (d) How old is your home? Pre First World War?
Interwar?
Post war?
Brand new?
- (e) Do you have exclusive use of bathroom and kitchen facilities?
- (f) How would you describe the area in which you live?
(look for key words)

(ii) Accessibility to health care

- (a) How far is it to your G.P.?
- (b) How long does it take you to get there?
- (c) What form of transport do you use to get there?
- (d) Is your G.P. the nearest to you?
- (e) If not, why not? (eg. nearest has full list, kept G.P. from before a move etc.)
- (f) Do you attend an antenatal clinic at your G.P.'s surgery?
(If not, repeat questions (a) to (e) for the clinic)
- (g) If your G.P./antenatal clinic were nearer to you, would you attend more frequently?

(iii) Health care provision

- (a) Were you satisfied with your medical treatment during pregnancy?
If not, why not?
- (b) Where did you learn most about your pregnancy?
from your G.P.?
from your midwife?
antenatal classes?
posters/N.H.S. literature?
friends/relations?
other?

(iv) Socio-economic environment

- (a) Is your partner unemployed?
- (b) If you were working during your pregnancy do you intend to return to work as soon as possible?
- (c) At what age did you finish full-time education?
(16?, 18?, 21?, other?)

(v) Maternal

Assessment of:

- (a) 'Wantedness' of the pregnancy
- (b) Pre-pregnancy weight, and maternal weight gain in pregnancy

Births Records are included in the project in order to standardise the mortality data and to assess the representativeness of the control population. It is recognised however, that the very limited range of variables included in this data set make it less than ideal. One potentially very interesting variable, the antenatal screening clinic attended by the mother, was rendered useless by problems of inconsistent computer coding.

Problems of confidentiality, political sensitivity and computer incompatibility, ruled out several potentially very useful data sources. Local child health and immunisation records would have helped to assess health problems beyond the perinatal period; infant and child mortality records could have further extended this work; and details of rateable values, rent and rate rebates and magistrates courts records (all in existence locally) could have been used to build a more meaningful social and economic picture of the study area.

The research into classifying census areas represents an important and successful contribution to the description of the demographic and socio-economic character of this urban area. Enumeration districts for the contiguous built-up area are classified into 9 broadly similar categories. An extension of this work could be to produce a 'purposive' classification, based on either the perinatal data, or the alternative census variables, or perhaps both. Conceptual problems however, would make a joint classification difficult since the census data pertain to areas, whilst the perinatal data refer to people. Allocating people to areas is a more acceptable practice in social analysis than assuming areal characteristics apply to individuals but the relatively few observations in the perinatal survey and their non-uniform spread across the study area would require careful interpretation of any resulting classification. A classification

based entirely on the survey data would of course be one of people, not areas, and would not have been ideally suited to the purposes of an investigation of geographical patterns and relationships.

Possibly the most fruitful type of purposive classification would be one based on an alternative set of census variables such as that described in Table 8.5. In it, each variable listed refers to the ratio of young children living in 'deprived' circumstances, such as the proportion of people aged 0-4 living at a density of greater than one person per room. A classification of areas according to these variables would identify zones of greatest deprivation among young children. However, it is doubtful if this would have added substantively to the results of the more general classification used.

Having reduced the original 847 EDs in Greater Leicester into 227 new small areas (NSAs) it was then possible to use SASPAC to extract a further set of census variables and to plot values for these using CHORO. Since each NSA comprised one or more census EDs the 'NEW ZONE' command within SASPAC was used and values were automatically aggregated for the constituent EDs. A constraint on this procedure is the amount of computing power required. This meant that in this study it took nine separate runs (one per cluster) to extract the necessary data. This is typical of the 'large data array' problems that were repeatedly encountered throughout this project. Although computer resources were both available and well-documented and therefore theoretically easy to use, in practice processing limits (both of time and memory) were continually hit. This necessitated the use and development of computer packages or options which had not previously been implemented at Leicester.

In the descriptive 'atlas' of Greater Leicester (Chapter 6) each of the census variables is mapped in turn, showing patterns of demography, ethnicity, socio-economic status and housing. This provides very useful background information, notwithstanding the simplicity and relative subjectivity of the map descriptions. The main disadvantage is the problem of how to establish direct comparability with the perinatal data.

Complementing this study of census variables is an analysis of perinatal data. This involves a case-control comparison and an attempt to identify 'areas of risk' using the distribution of survey members exhibiting risk characteristics. A total of twenty variables are found to be associated with perinatal death (see Table 7.26), most significant of these are various infant characteristics (such as low birthweight and low gestational age) and health care variables (including measures of attendance for antenatal care). Some physical, obstetric and socio-economic characteristics are also significantly associated with adverse pregnancy outcome.

By introducing the Leicestershire Births data into this analysis it is possible both to examine the overall spatial distribution of births in Greater Leicester, and tentatively to confirm the representativeness of the control population. Of the six variables common to both births and survey data, the control population correctly identifies the direction of the association in five, the exception being maternal age which only exhibits a significant association when cases are compared with all births.

In the section relating the distribution of perinatal death to that of 'risk' characteristics, postcode coordinates are firstly obtained for each of the perinatal survey members. These are then used as a basis for assigning ED codes and cluster names to the survey data. The coding convention for EDs means that ward

and district names are implicit from the ED names. Two different analyses are then employed. The first is a statistical analysis of the distribution of perinatal deaths by ward, district and cluster, and the second a cartographic analysis which allows for a spatial standardisation of the births data. Of the two analyses the former, which compares mortalities and total births, yields the most significant result, particularly when cluster membership is examined. The cluster comprising areas dominated by the low status and ethnic minority populations has the highest perinatal death rates. This result is confirmed both by the general maps showing perinatal death rates and the plotted distributions of mothers exhibiting risk characteristics. Several of the 'risk' factors are classes of socio-economic variables (for example the 'low status' socio-economic groups, the Indian and West Indian ethnic groups and the African and Asian regions of birth) which are spatially concentrated, especially around the central city areas of Highfields and Belgrave.

If time had been available, it would have been particularly interesting to have extended this analysis to include deaths in the post-neonatal period, in which the young baby is more directly exposed to the home social and economic environment. Although death rates in this period are much lower, a significant association might still have been hypothesized.

The attempt to bring together and extend the census variable and perinatal survey analysis involves several methodologies. First, a principal components analysis is performed, using the original set of 34 census variables for the 227 NSAs created by the classification. The resulting component scores are then mapped and the maps compared with a map of the perinatal mortality rate. Unfortunately the interpretation both of the components and of the

maps is rendered difficult by a set of weak component loadings and some uncertainty as to the appropriateness of using these data in this technique. In any case no significant association is found between any of the components identified and perinatal mortality.

The second method involves the calculation of areal association using a correlation coefficient. Three types of mapped index (absolute, ratio and chi-square) are compared with perinatal mortalities. Other than a spurious positive relationship brought about by a third variable, area size, no significant associations are found.

The calculation of a 'risk surface' is the third attempt at using census data to explain the pattern of perinatal deaths. An 'observed' surface is calculated by assigning mortalities and births to grid squares, producing a data set which could then be plotted by the GHOST80 graphics package. Several 'risk surfaces' based on different 'risk indexes' computed from the census and perinatal data are then calculated. This is done in order to identify those risk indexes which can predict the actual pattern of mortalities. Unfortunately this analysis could not be successfully completed owing to the irregular distribution of mortalities and the large number of grid squares with zero values.

While the calculation of the risk index involves the subjective assessment of the relative importance of the different risk factors, the subsequent application of stepwise logistic regression to the perinatal data quantifies the relationships among the variables. Of all the models tested, only one satisfactorily predicts the perinatal outcome. In order of significance the variables entered were as follows:

- (i) length of pregnancy
- (ii) infant birthweight
- (iii) type of labour
- (iv) gravid state
- (v) father's socio-economic group

- (vi) previous perinatal death
- (vii) number of days in-patient antenatal care
- (viii) availability of an antenatal clinic session at the G.P.'s surgery.

Subsequent attempts to assess the joint effects of combinations of variables failed to produce significant results, probably because there were insufficient observations for the number of possible combinations of variables and categories. Indeed, this was the major limiting factor in the use of the log-linear method. Where geographers have previously applied such techniques, they have usually considered only two or three variables, each of no more than four or five categories. Perinatal mortality is too complex to be explained by so few factors, even if the variables are reduced to a simple risk/no risk categorisation; the overall number of data cells in such a model will still be 2^n (where n is the number of variables).

The final procedure employed is the calculation and plotting of Poisson probabilities of perinatal death. Based on a comparison of actual and expected numbers of mortalities, the probability map highlights areas with particularly good or severe mortality experience. Four maps are plotted, each representing a different measure of the 'at risk' population. All four show the relatively 'deprived' areas of Highfields, Belgrave, the inner city and the Aylestone Road to have perinatal mortality rates higher than expected. The maps confirm the project's hypothesis that areas of social and economic disadvantage also experience the worst perinatal outcome.

9.2 Conclusions

In addition to the new knowledge of perinatal mortality provided by the analysis of the different data sets, this research contributes to the geographer's understanding of an urban health problem in various ways. In the first instance, literature from

both medicine and the other social sciences has been examined and reviewed, constituting a truly interdisciplinary approach. Secondly, the interdisciplinary theme has been continued by collaboration with local health care providers and planners, and through these, policy makers. This thesis is offered to the latter for consideration in planning future health care. It is significant that the geographical perspective has previously been lacking in studies of perinatal mortality, despite the clear spatial implications identified by this project. Although limited by the format of the data sources used, this thesis has suggested a full geographical interpretation of these data.

With respect to the methodologies employed, this study makes several contributions to geographical research. The application of every technique has been a means to an end, with justification given for each stage of the analysis. The major sequence of techniques used in classifying census EDs has universal applicability, either part or all of the different steps could be usefully adopted by other social scientists, indeed the framework of this thesis could serve as a foundation for a variety of similar projects.

The use of multiple data sets in this work offers new and exciting possibilities for social research. Both the volume of data handled here and the problems overcome in combining spatially inconsistent data sets represent major contributions to perinatal studies.

9.3 Suggestions for further research

A study of this nature inevitably raises almost as many questions as it answers. This project has provided a new insight into the perinatal problem in Greater Leicester and, by analogy, some indication of what might be expected in other urban areas. It would be gratifying to think that other researchers could adopt

some of the methods used here for studies in other cities or of other health problems.

Having completed this work, several suggestions for further research can be made. In particular, aspects of health care during pregnancy would bear further examination. The distribution of health services, accessibility to and uptake of care, attitudes to care, health education, the impact of NHS policy changes, and the assessment of the quality of care might all be considered. Changes in health care, perinatal mortality and urban structure all have a temporal dimension which has not been examined here. A comparison of the 1981 census results with those of 1971 or 1961 would show some significant changes, particularly in the ethnic structure of the city's population. Assuming that these changes also have a spatial dimension, any change in the distribution of perinatal mortality might also be reassessed.

An obvious extension to this particular project would be the further investigation of those areas identified as having the worst perinatal experience. A further survey into the behaviour and attitudes of maternity patients in these areas would be of special interest, as would a similar survey of local health care providers. It is increasingly apparent that the greatest short term benefits will arise from direct action on the part of these two groups, even if in the long term significant improvements in perinatal health will occur only as a result of an improvement in the social, economic and health status of the entire population.

Appendix 1. Mapping Census Data

A1.1 Introduction

This appendix describes the results of a mapping exercise, the objectives of which were to examine and display the effects of various operational decisions on a computer-drawn map. The transformation of variables, style of map and choice of class intervals are all considered and, following an evaluation of the different maps, recommendations are made for the plotting of the remainder of the census variables.

Three census variables were chosen as being representative of the majority of variables to be plotted. In each case ratio values were initially used. The first, 'Residents aged 45 to retirement' has a central, unimodal distribution (see Figure A1.1), the second, 'Households in owner-occupied accommodation' has a U-shaped distribution, and the third, 'Households in privately rented property' is highly negatively skewed. Given the area-based nature of the census data the choropleth style of map was felt to be appropriate. The use of the choropleth map for displaying ratio-type census variables is common in the literature (see for example Maguire et al., 1983) and is an appropriate starting point for the mapping exercise.

A1.2 Choice of Class Intervals

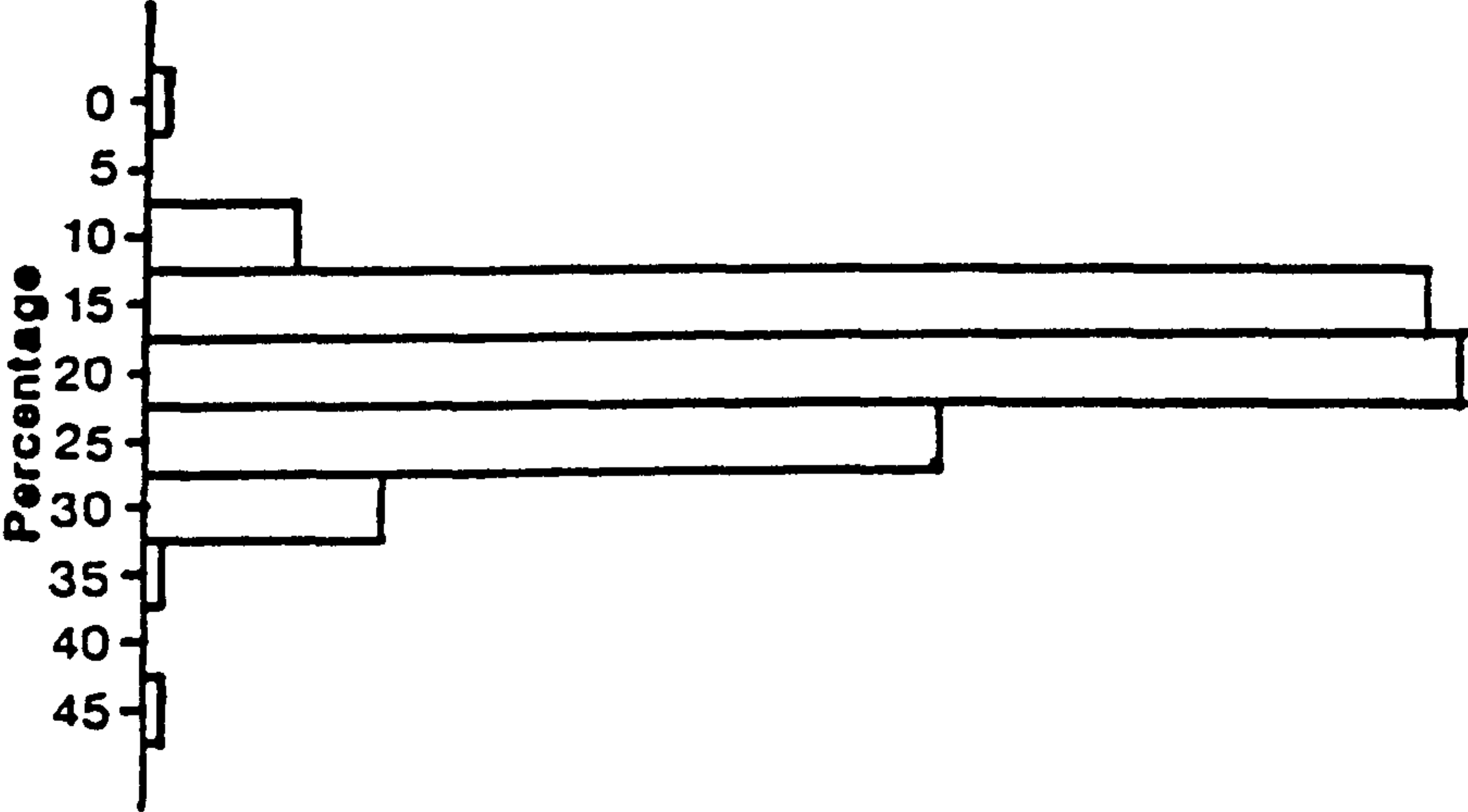
The first decisions in choropleth mapping are whether to define classes or not (see Tobler, 1973 for a discussion of mapping without class intervals) and, if so, how many and what limits the classes should have.

Jenks and Coulson (1963) recommend that class intervals should

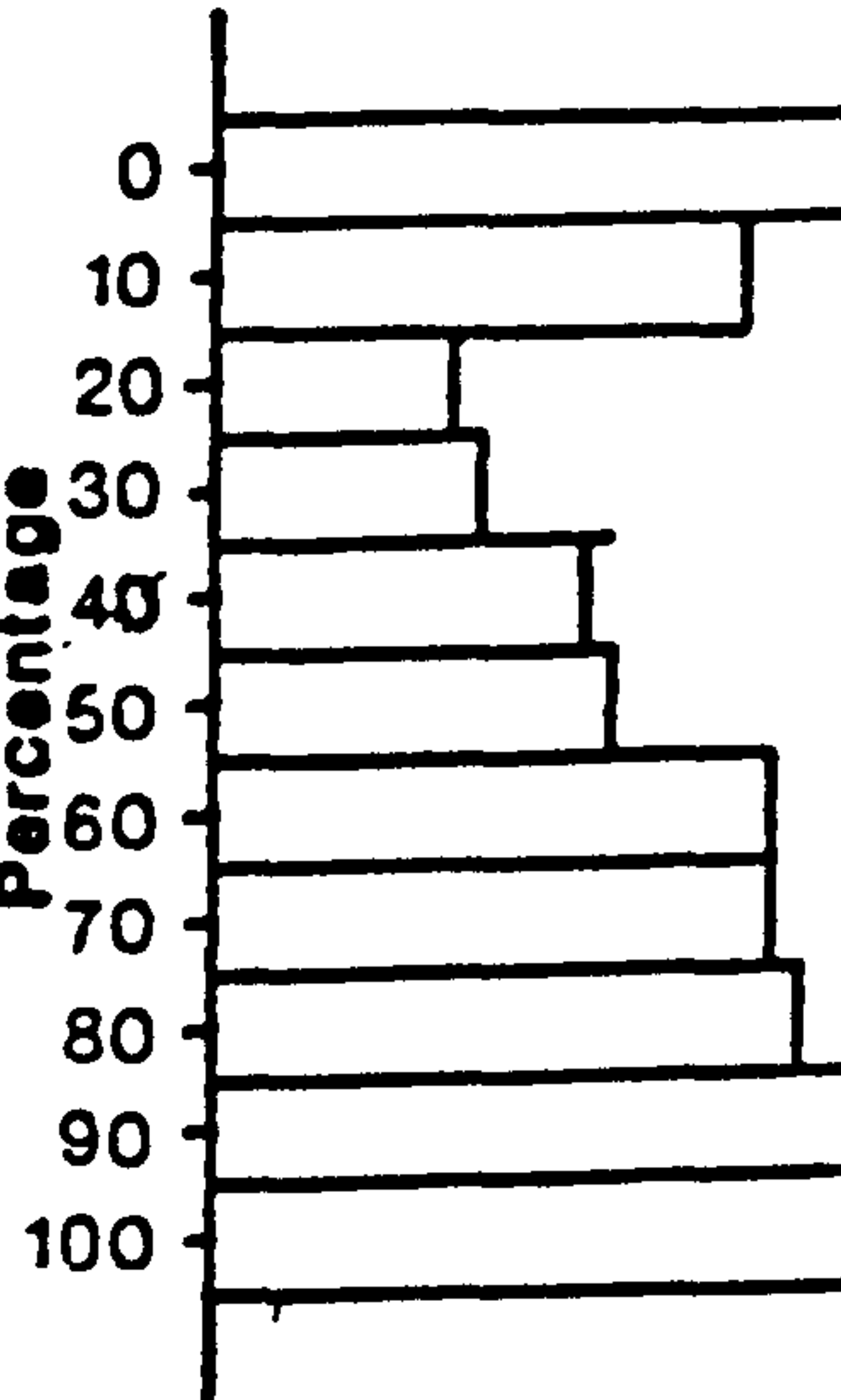
- (i) encompass the full range of the data
- (ii) have neither overlapping values nor vacant classes
- (iii) be great enough in number to avoid sacrificing the accuracy

Figure A1.1 Histograms of selected census variables

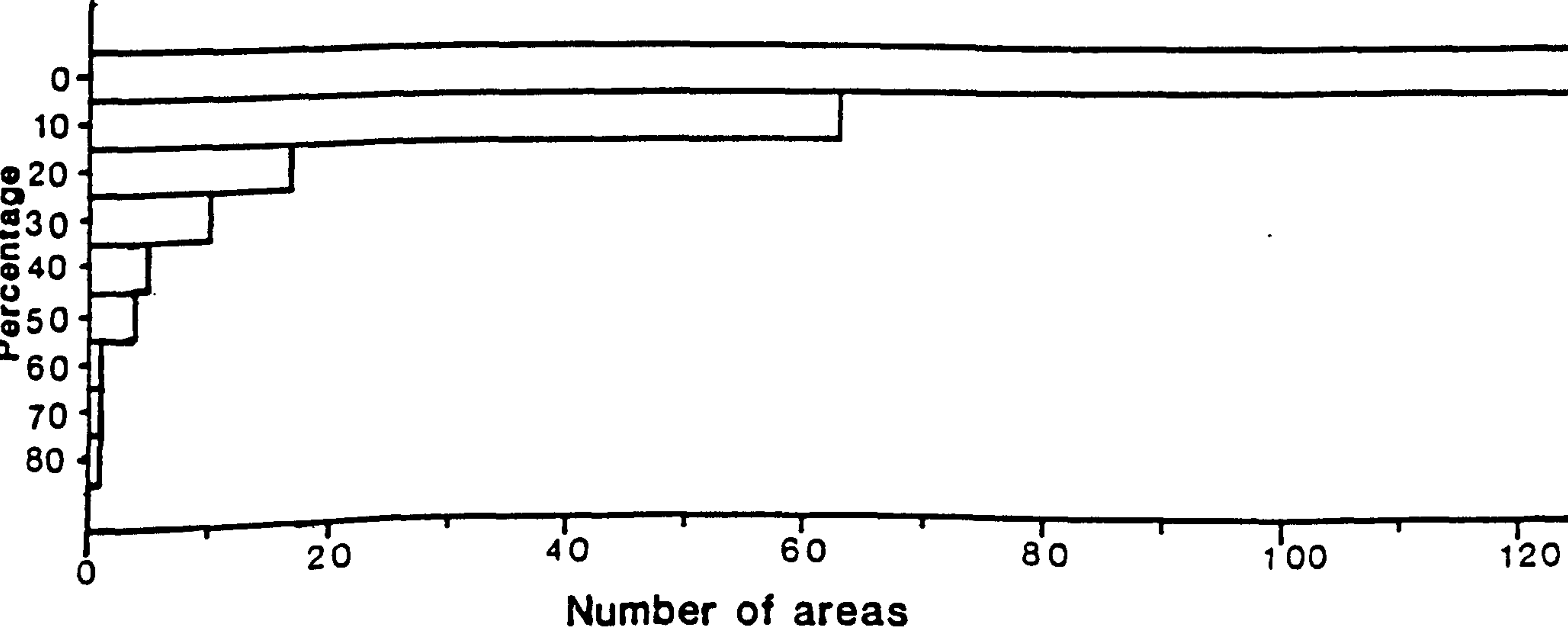
Residents aged 45 to retirement



Households in owner-occupied accommodation



Households in privately rented property



of the data, but not be so numerous as to input a greater degree of accuracy than is warranted by the nature of the collected observations

(iv) divide the data into reasonably equal groups of observations

(v) have a logical mathematical relationship if practicable

Above all, the intervals should be appropriate to the statistical frequency distribution of the data.

Evans (1977) suggests that there are four approaches to class interval selection. First there is the 'exogenous' approach, in which intervals are chosen at meaningful threshold values relevant to the data being mapped. An example is a critical population density value such as 1.5 persons per room in an urban area. The second approach is the 'arbitrary' one involving round numbers that have no particular significance. Interval size need not be constant, for example the set of class boundaries 5, 10, 20, 30, 40, 80, 120. The 'idiographic' approach is the third approach, in this, intervals are based on specific characteristics of the data set being mapped. Multimodal, multistep, contiguity biased, correlation biased, percentile and nested mean class limits are all examples of idiographic methods. Finally there is the 'serial' approach in which class limits have a definite mathematical relation to each other. Evans suggests various arithmetic, geometric and curvilinear progressions, equal intervals based on different scales, and a system based on standard deviations from the mean.

Of these possibilities, Rhind (1983) suggests that four might be appropriate for mapping census data: the 'important thresholds' method, the 'multimodal' method (which he does not recommend), the arithmetic or geometric progression, and the mean and standard deviation approach.

Based on these recommendations it was decided that six methods of class interval selection would be tested. These are outlined below.

- (i) **'Multimodal' or 'break points'**. Dispersion diagrams are examined for natural 'breaks' in the variable distributions, thus both the number of classes and the class limits are determined by the data.
- (ii) **Arithmetic progression**. With the exception of the highest and lowest classes, each class is of the same width and class limits are defined by 'round' numbers.
- (iii) **Geometric progression**. This method follows the same principle as the arithmetic progression but class intervals progressively increase, for example having limits of 0, 2, 6, 14, 30, 62.... A geometric progression is especially suited to data which are skewed towards 0.
- (iv) **Percentiles**. Percentile classes may contain either equal numbers of spatial divisions or near-equal areas (Evans, 1977), where the variables plotted measure an unevenly distributed population the former is more appropriate. In this experiment data values were ranked, the number of classes decided upon, and classes filled by equal numbers of areas from the ranked set. This method is useful for widely and evenly spaced data values, but insensitive at identifying extreme values.
- (v) **Mean and standard deviations**. The standard deviation is a measure of spread within a distribution. Class limits are based on the mean value and calculated as a proportion of the standard deviation either side. It is preferable to ensure that the mean value lies at the centre of a class (which will result in an odd number of classes) rather than determining a class limit (and resulting in an even number

of classes).

(vi) **Nested means.** Scriptor (1970) suggests this to be an alternative to the standard deviation for defining classes about a mean value. A first order mean is calculated for the entire data set, followed by second order means for those values above and below the first order mean, third order means in between these, and so forth. This method limits the user to 2, 4, 8, 16, ..., 2^n classes.

The **important thresholds** method was considered to be potentially useful but inappropriate to the variables being assessed.

Table A1.1 shows the class limits calculated for each of the variables and the methods of selection.

A1.3 Comparing class interval systems

Assessment of the usefulness of each set of class intervals is inevitably a subjective procedure. In order to reduce this subjectivity the following criteria were adopted as standards against which the maps might be judged:

- (i) comparability with maps of other census variables
- (ii) class intervals to be meaningful in the context of the data set
- (iii) readability - different classes to be clearly identifiable
- (iv) aesthetic value - the general appearance of the map to be as 'balanced' as the data allow
- (v) ability to distinguish exceptionally high and low values in the data limits.

Figures A1.2 to A1.19 show the effects of changing class limits.

Considering first the variable 'Residents aged 45 to retirement', for these data the most useful display is that based on percentiles (Figure A1.2). This method successfully identifies

Table A1.1 Class intervals for map production

1. Residents aged 45 to retirement	2. Households in owner-occupied accommodation	3. Households in privately rented property

MULTIMODAL (BREAK POINTS)		
1 0.019 to 0.14	0.0 to 0.11	0.0 to 0.01
2 0.141 to 0.19	0.111 to 0.61	0.011 to 0.04
3 0.191 to 0.22	0.611 to 0.87	0.041 to 0.08
4 0.221 to 0.27	0.871 to 1.0	0.081 to 0.13
5 0.271 to 0.439		0.131 to 0.79
ARITHMETIC PROGRESSION		
1 0.019 to 0.05	0.0 to 0.2	0.0 to 0.1
2 0.051 to 0.15	0.21 to 0.4	0.11 to 0.2
3 0.151 to 0.25	0.41 to 0.6	0.21 to 0.3
4 0.251 to 0.35	0.61 to 0.8	0.31 to 0.4
5 0.351 to 0.439	0.81 to 1.0	0.41 to 0.79
GEOMETRIC PROGRESSION		
1 0.019 to 0.02	0.0 to 0.05	0.0 to 0.03
2 0.021 to 0.06	0.051 to 0.15	0.031 to 0.09
3 0.061 to 0.14	0.151 to 0.35	0.091 to 0.21
4 0.141 to 0.30	0.351 to 0.75	0.211 to 0.45
5 0.301 to 0.439	0.751 to 1.0	0.451 to 0.79
PERCENTILES		
1 0.019 to 0.152	0.0 to 0.185	0.0 to 0.01
2 0.153 to 0.177	0.186 to 0.537	0.011 to 0.029
3 0.178 to 0.206	0.538 to 0.758	0.03 to 0.058
4 0.207 to 0.236	0.759 to 0.903	0.059 to 0.108
5 0.237 to 0.439	0.904 to 1.0	0.109 to 0.79
MEAN AND STANDARD DEVIATIONS		
1 0.019 to 0.1195	0.0 to 0.0785	(no values)
2 0.1196 to 0.1705	0.0786 to 0.4095	0.0 to 0.024
3 0.1706 to 0.2215	0.4096 to 0.7405	0.025 to 0.152
4 0.2216 to 0.2725	0.7406 to 1.0	0.153 to 0.278
5 0.2726 to 0.439	(no values)	0.281 to 0.79
(\bar{x} =0.196, sd=0.051)	(\bar{x} =0.575, sd=0.331)	(\bar{x} =0.088, sd=0.128)
NESTED MEANS		
1 0.019 to 0.156	0.0 to 0.182	0.0 to 0.027
2 0.157 to 0.194	0.183 to 0.611	0.028 to 0.081
3 0.195 to 0.229	0.612 to 0.831	0.082 to 0.203
4 0.230 to 0.439	0.832 to 1.0	0.204 to 0.79

Note: 'Mean' values as shown for the 'mean and standard deviation' and 'nested mean' methods differ by a small percentage because they are products of different calculations. The nested means were calculated from absolute data values (i.e. total observations with a characteristic/total observations) before subsequent comparison with the ratio values, whilst the mean and standard deviations were of necessity calculated from the ratios themselves. This resulted in a value which was not a true mean since it failed to allow for the effects of different sized areas.

Figure A1.2

Comparing class interval systems for census variables

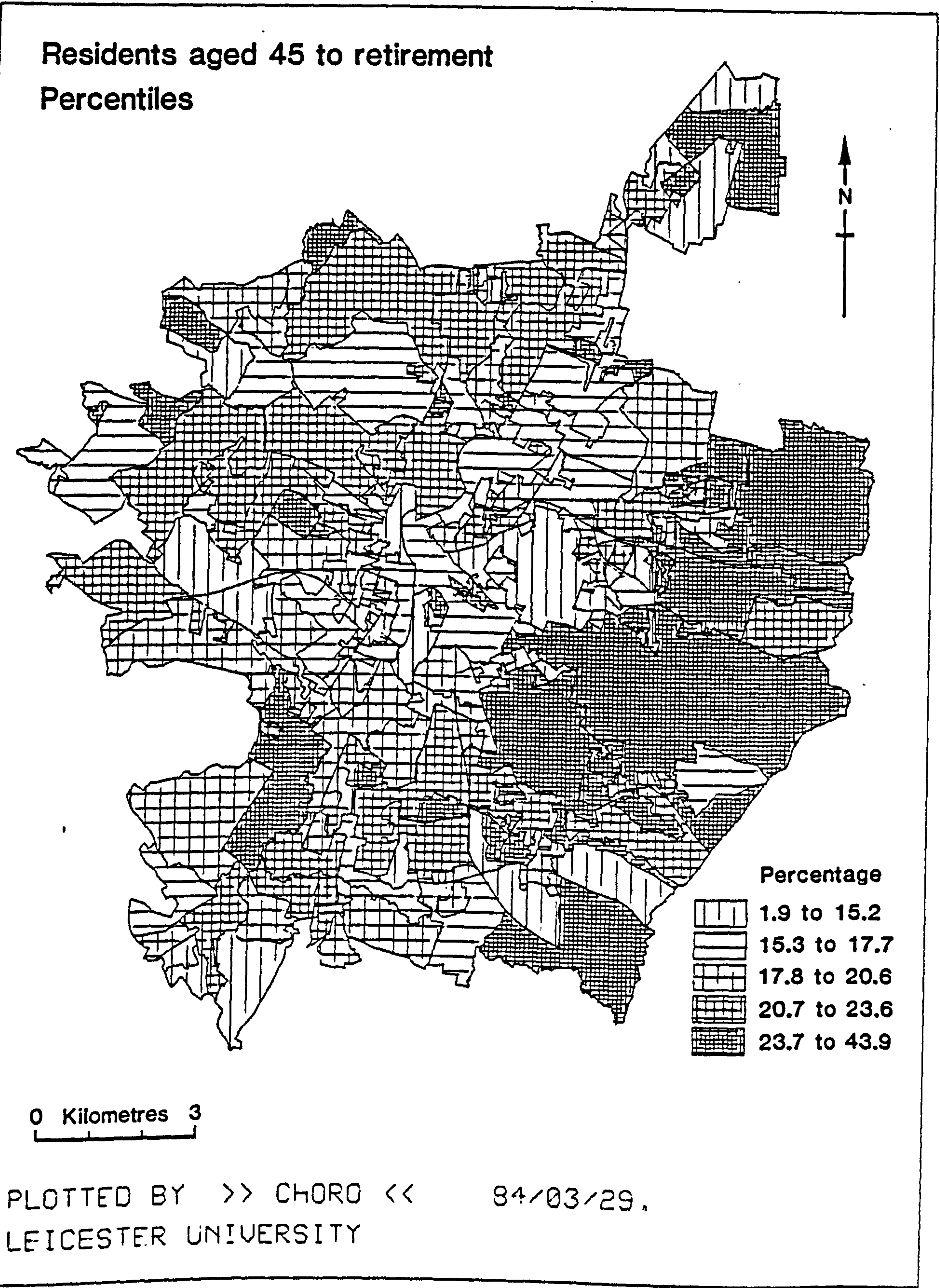
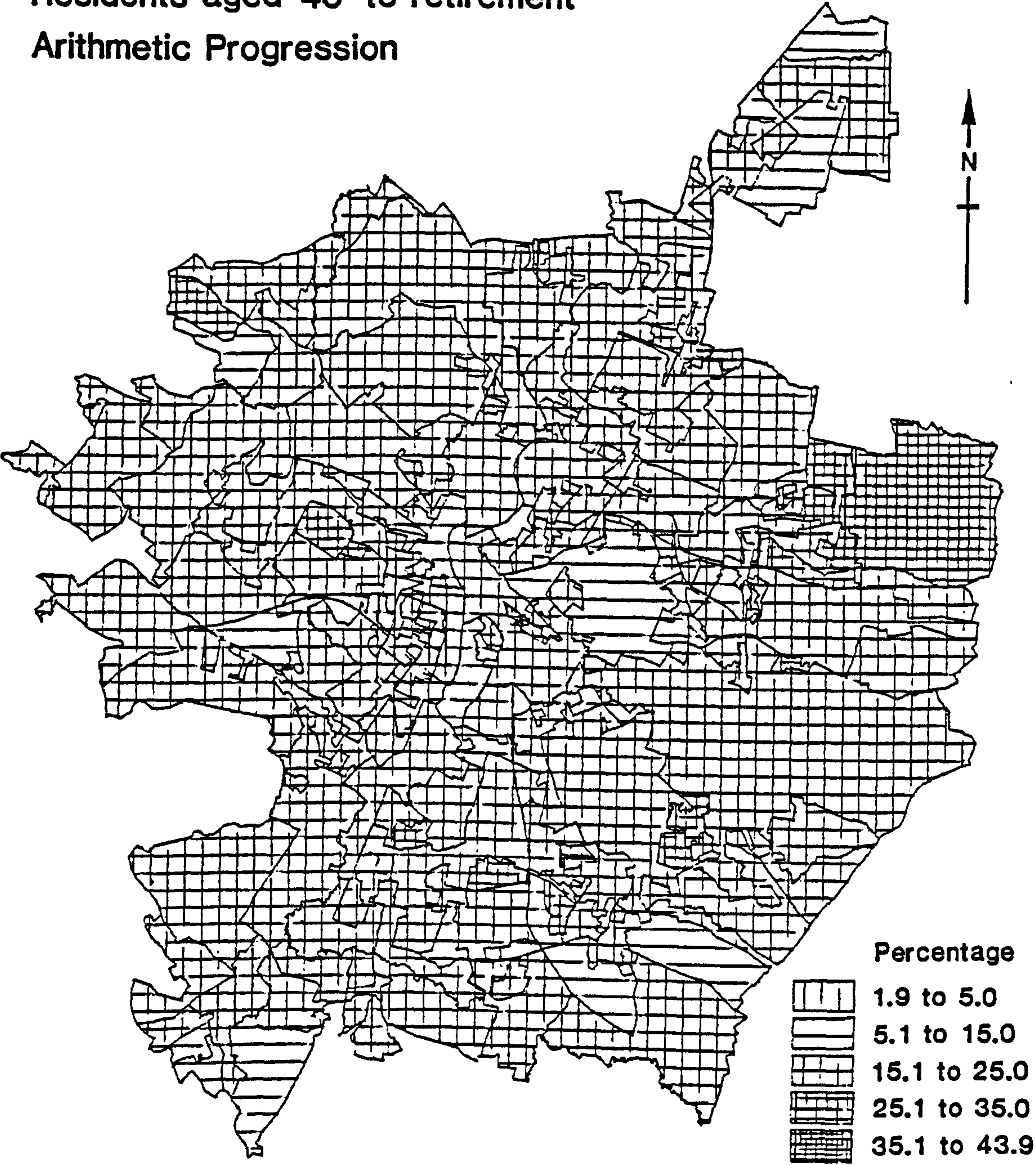


Figure A1.3

Residents aged 45 to retirement
Arithmetic Progression

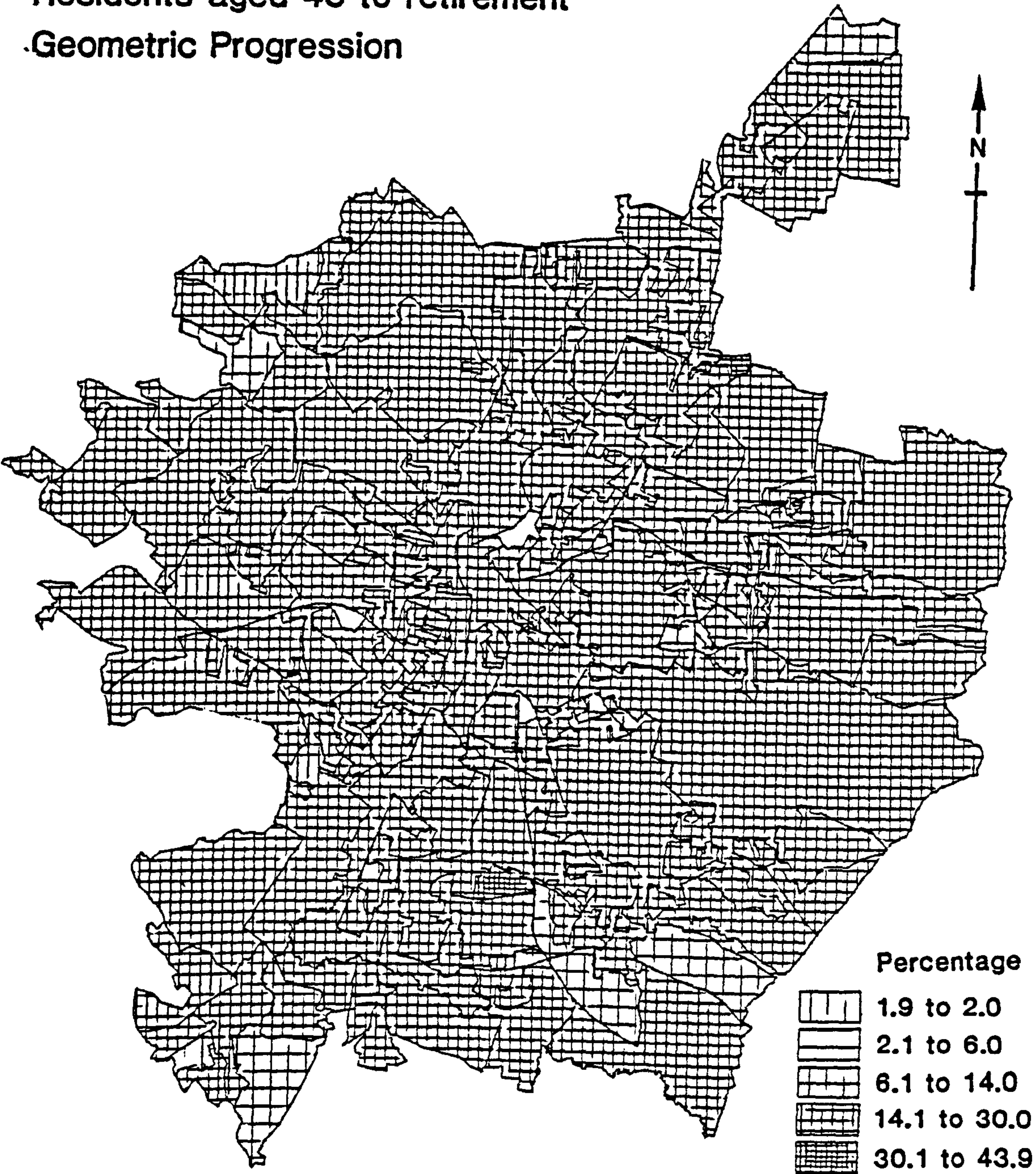


0 Kilometres 3

PLOTTED BY >> CHORO << 84/03/24.
LEICESTER UNIVERSITY

Figure A1.4

Residents aged 45 to retirement
Geometric Progression



0 Kilometres 3

PLOTTED BY >> CHORO << 84/03/24.
LEICESTER UNIVERSITY

Figure A1.5

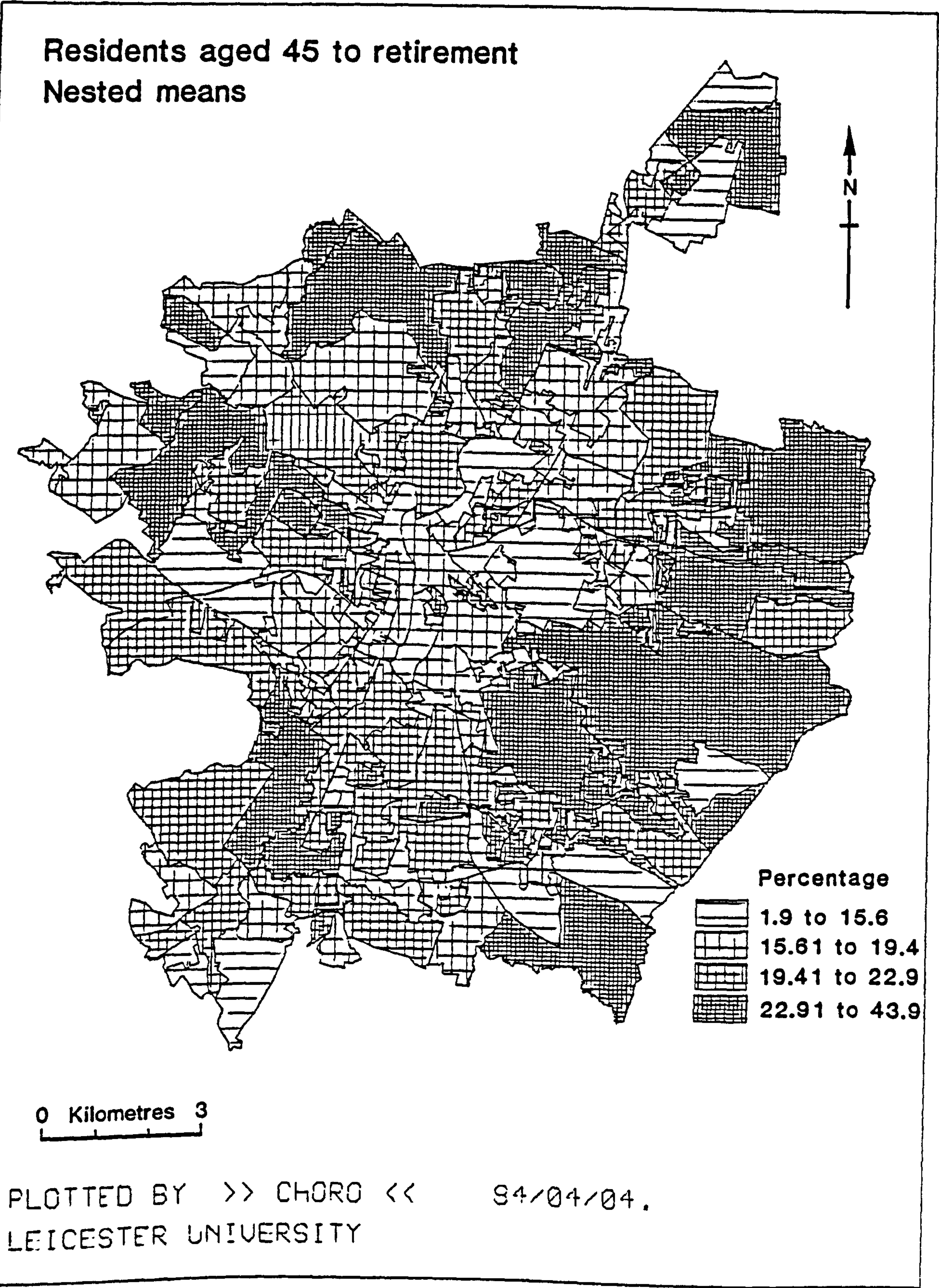


Figure A1.6

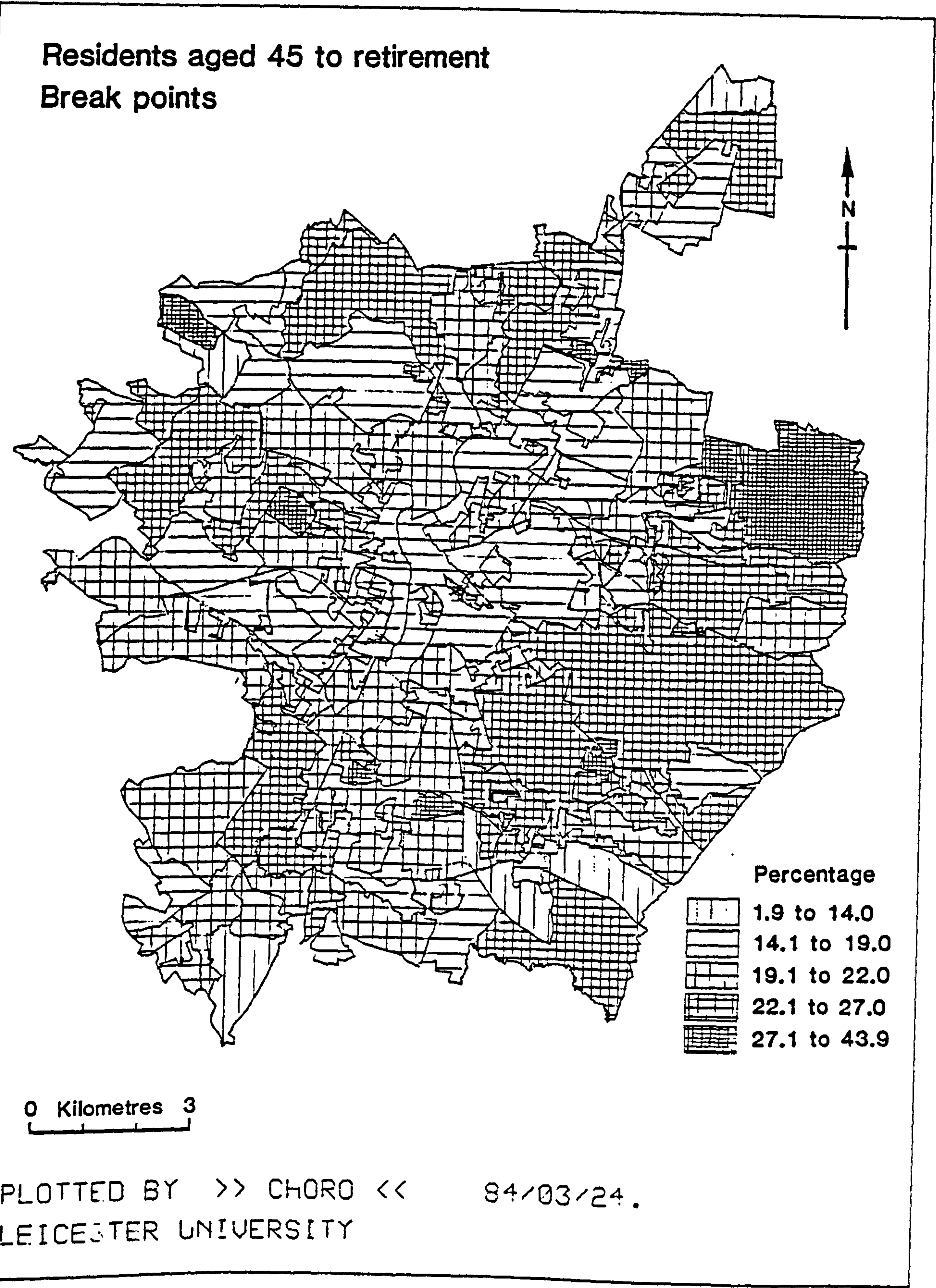
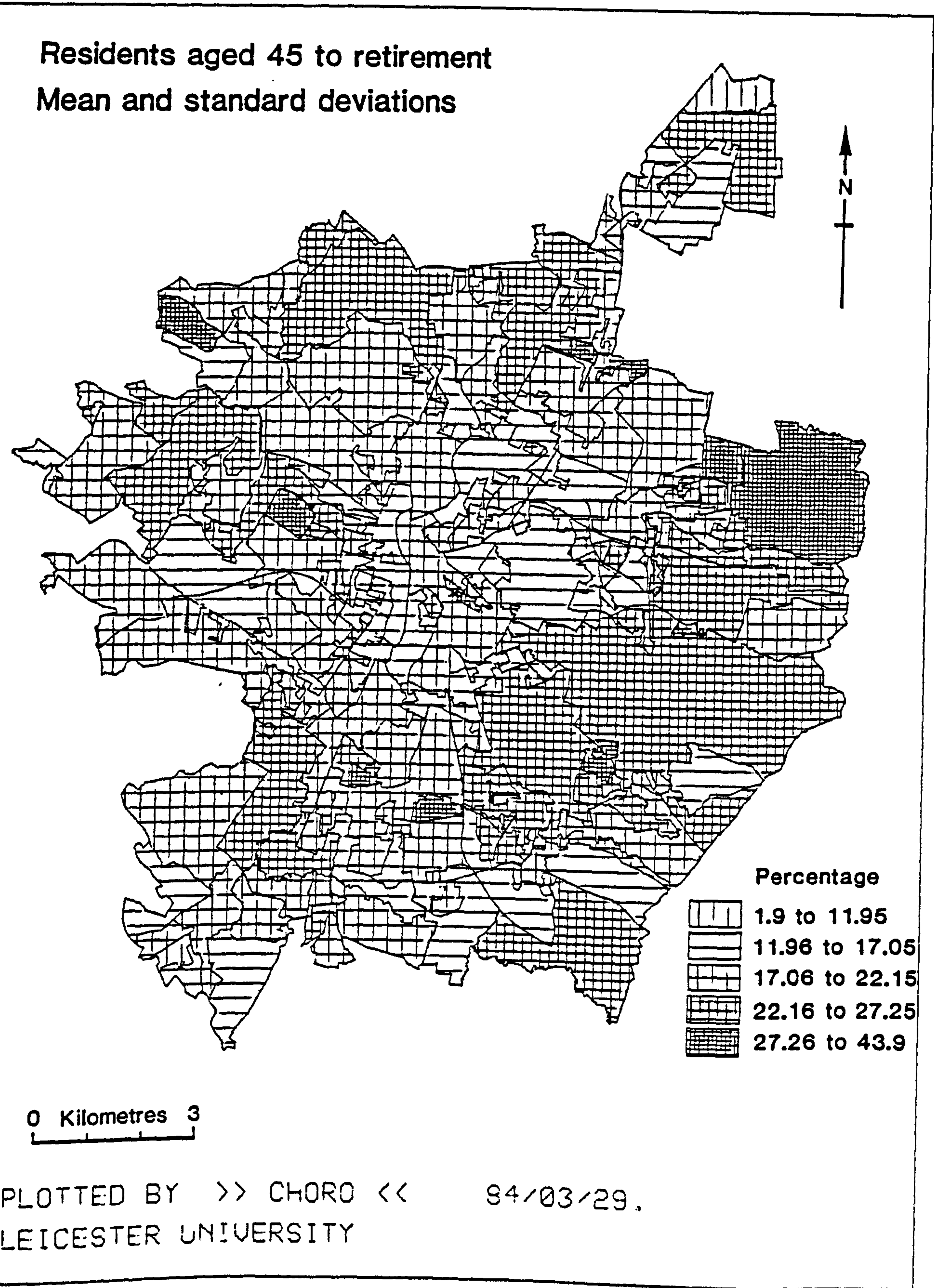


Figure A1.7



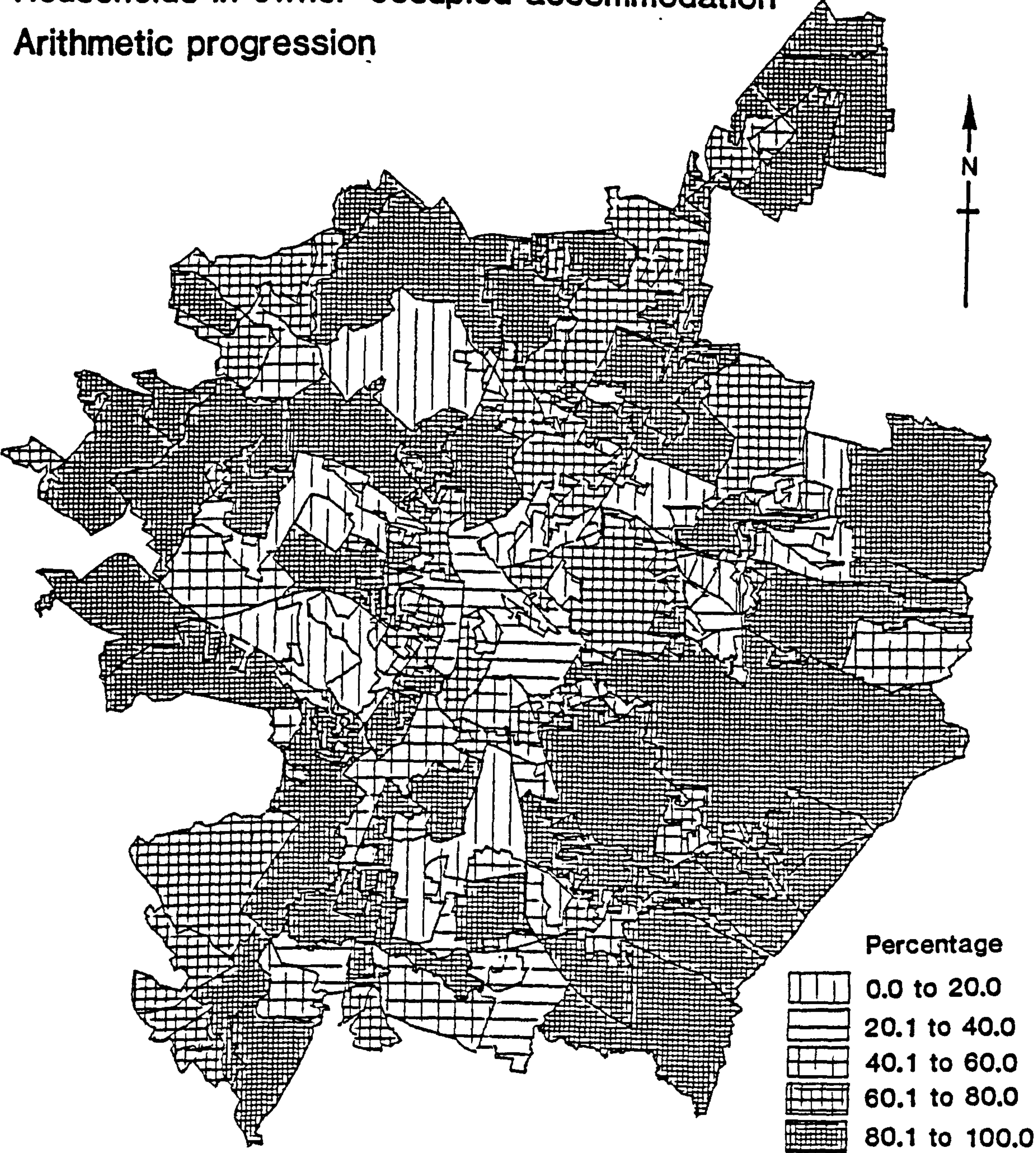
areas with high and low values and produces a clear and legible map. The least useful maps are produced from the arithmetic and geometric progressions (Figures A1.3 and A1.4), both of which have the majority of areas in a single category. Of the remaining maps, that based on nested means (Figure A1.5) suffers slightly compared with the percentiles map for having fewer classes, whilst the 'break points' (Figure A1.6) and mean and standard deviation systems (Figure A1.7) fail to satisfactorily identify extreme values.

The arithmetic progression is the most successful at displaying 'Households in owner-occupied accommodation' (Figure A1.8). Being conceptually the simplest system, the round numbers and regular intervals aid interpretation of the map and are particularly suited to closed number data sets. In contrast, the mean and standard deviation map (Figure A1.9) is unable adequately to represent the data because some of the calculated class limits lie beyond the range of values. The geometric progression (Figure A1.10) clearly identifies areas with low proportions of owner-occupied housing, but fails to distinguish areas lying at the other extreme of the scale. Again, this method produces class limits beyond the range of the data. There is a close similarity between the nested means (Figure A1.11) and 'break points' maps (Figure A1.12), this is due to there being a natural break coincident with the mean value. The percentiles map (Figure A1.13) correctly identifies the areas with highest and lowest data values, but completely disguises the underlying statistical distribution.

All six maps successfully highlight areas of high concentration of privately rented property, but the nested mean map (Figure A1.14) most satisfactorily identifies areas of low concentration. In this example fewer class intervals does not

Figure A1.8

Households in owner-occupied accommodation
Arithmetic progression

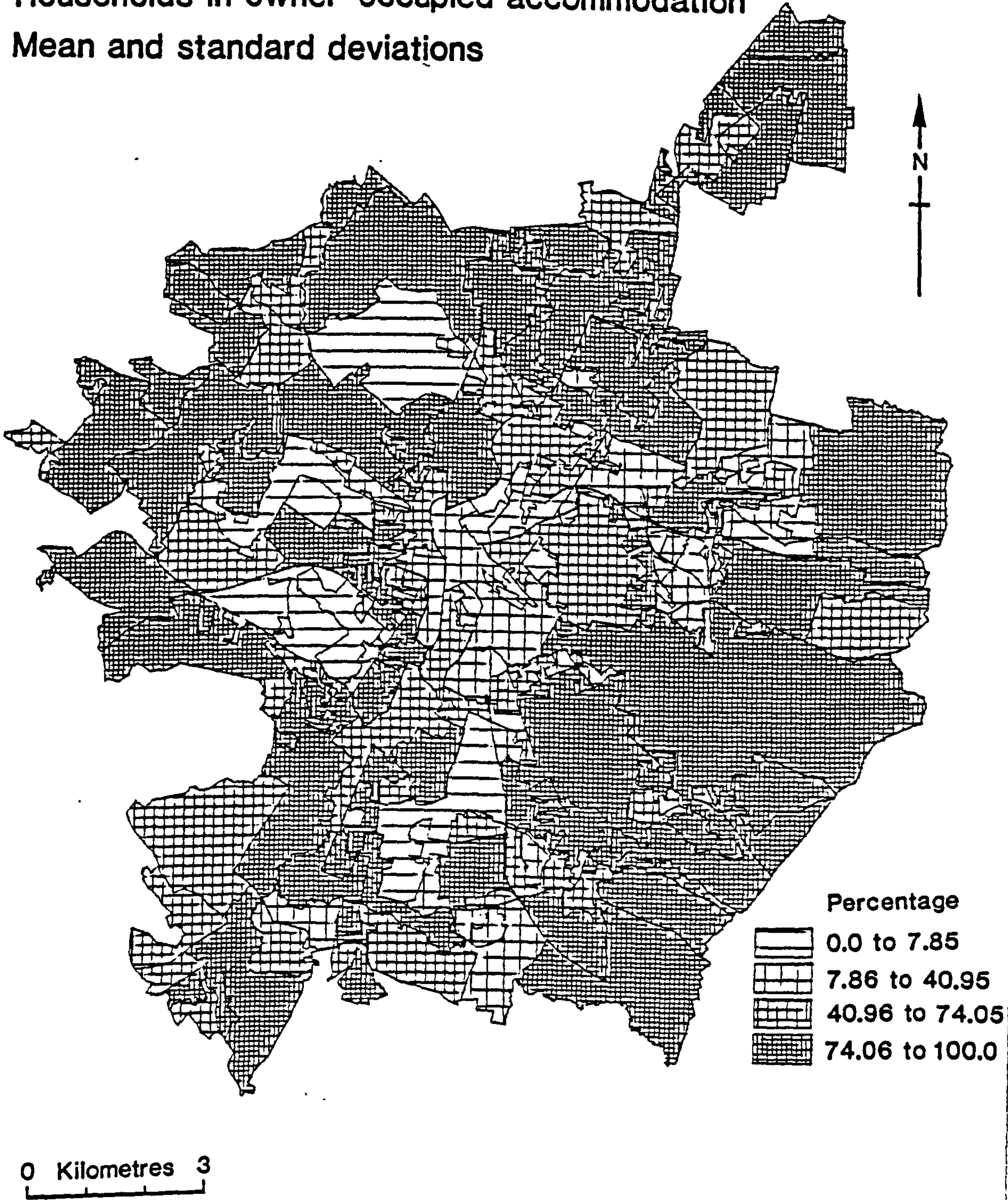


0 Kilometres 3

PLOTTED BY >> CHORO << 84/03/28.
LEICESTER UNIVERSITY

Figure A1.9

Households in owner-occupied accommodation
Mean and standard deviations



PLOTTED BY >> CHORO << 84/03/28.
LEICESTER UNIVERSITY

Figure A1.10

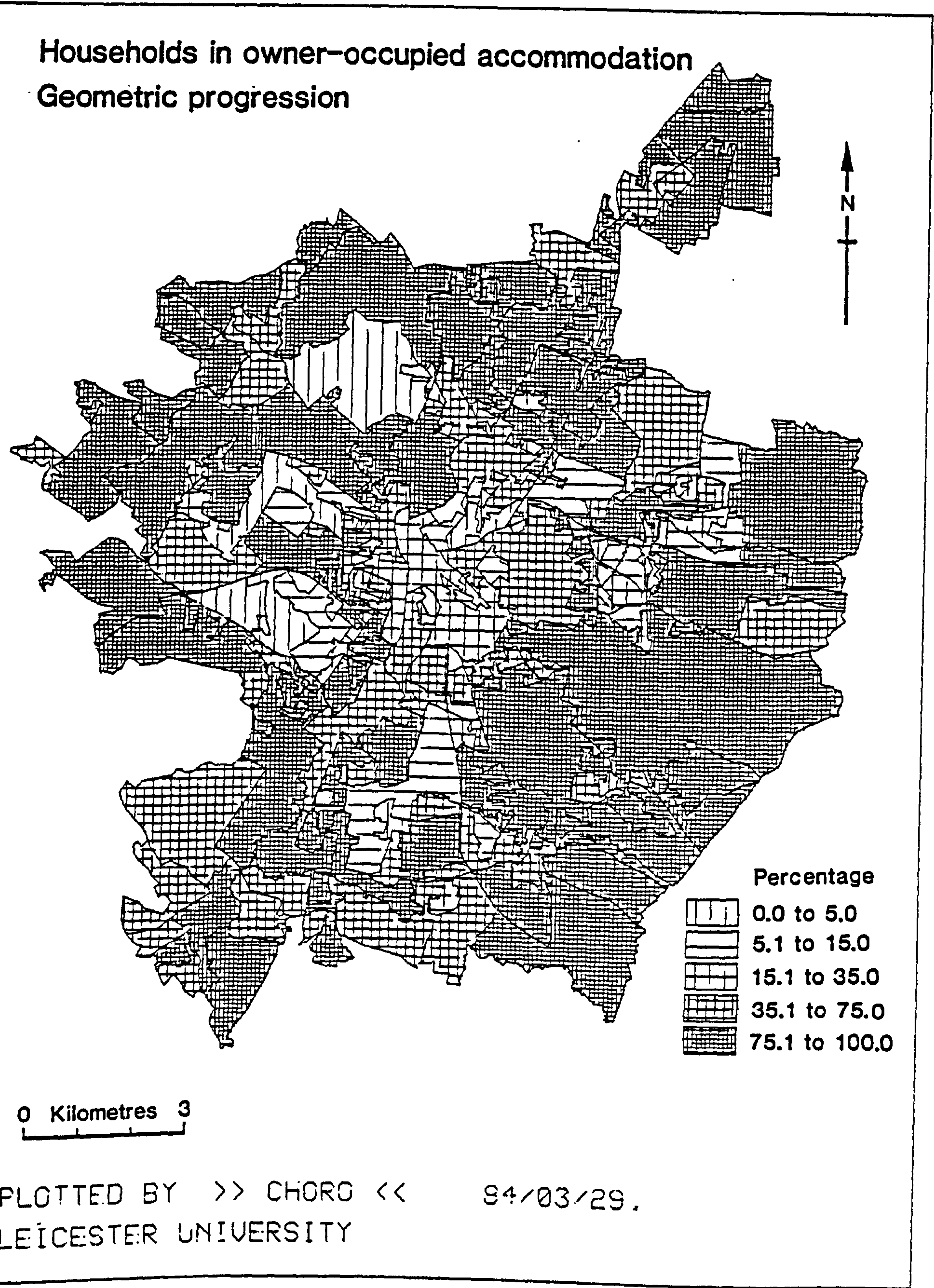
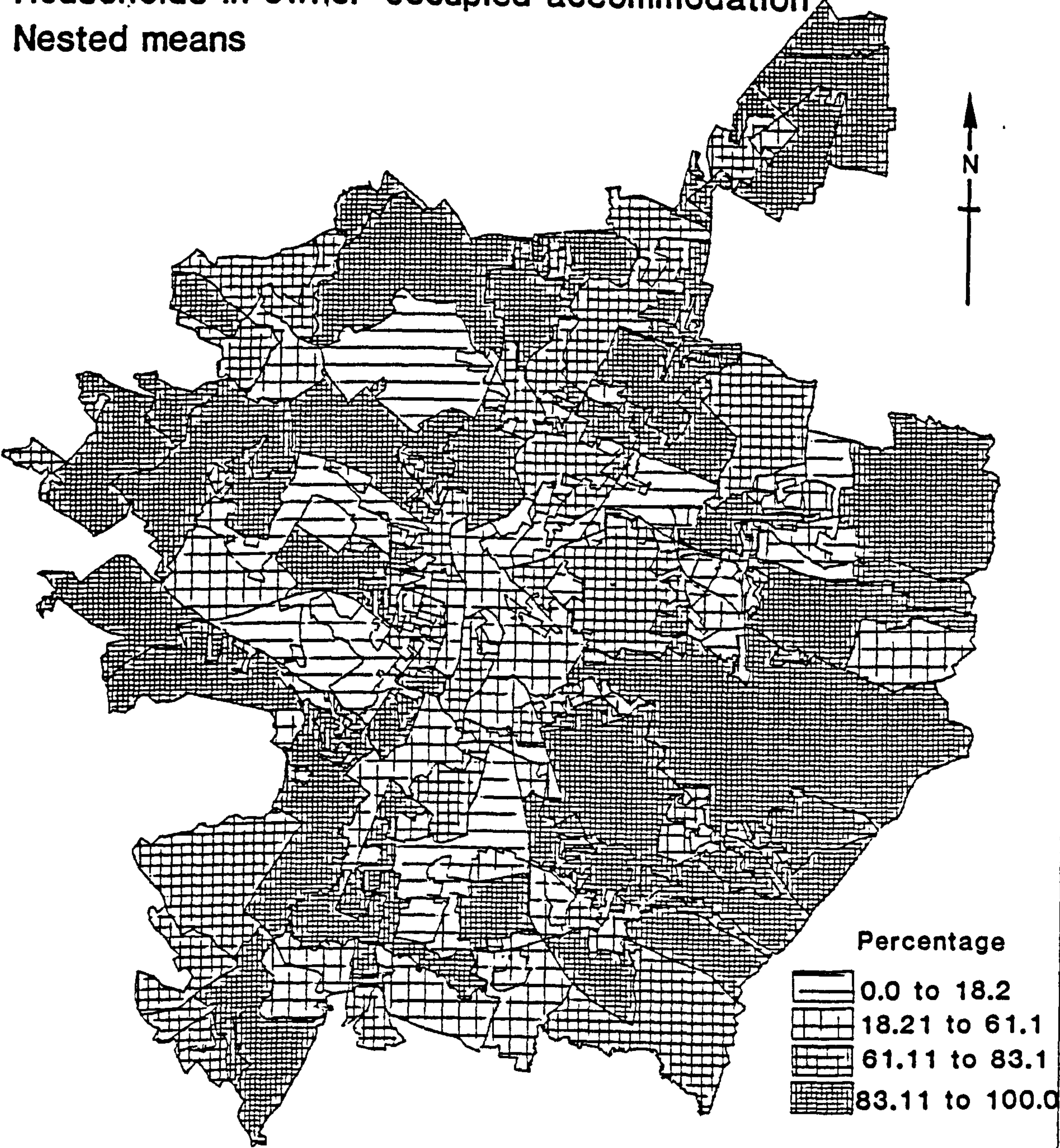


Figure A1.11

Households in owner-occupied accommodation
Nested means



0 Kilometres 3

PLOTTED BY >> CHORO << 84/04/04.
LEICESTER UNIVERSITY

Figure A1.12

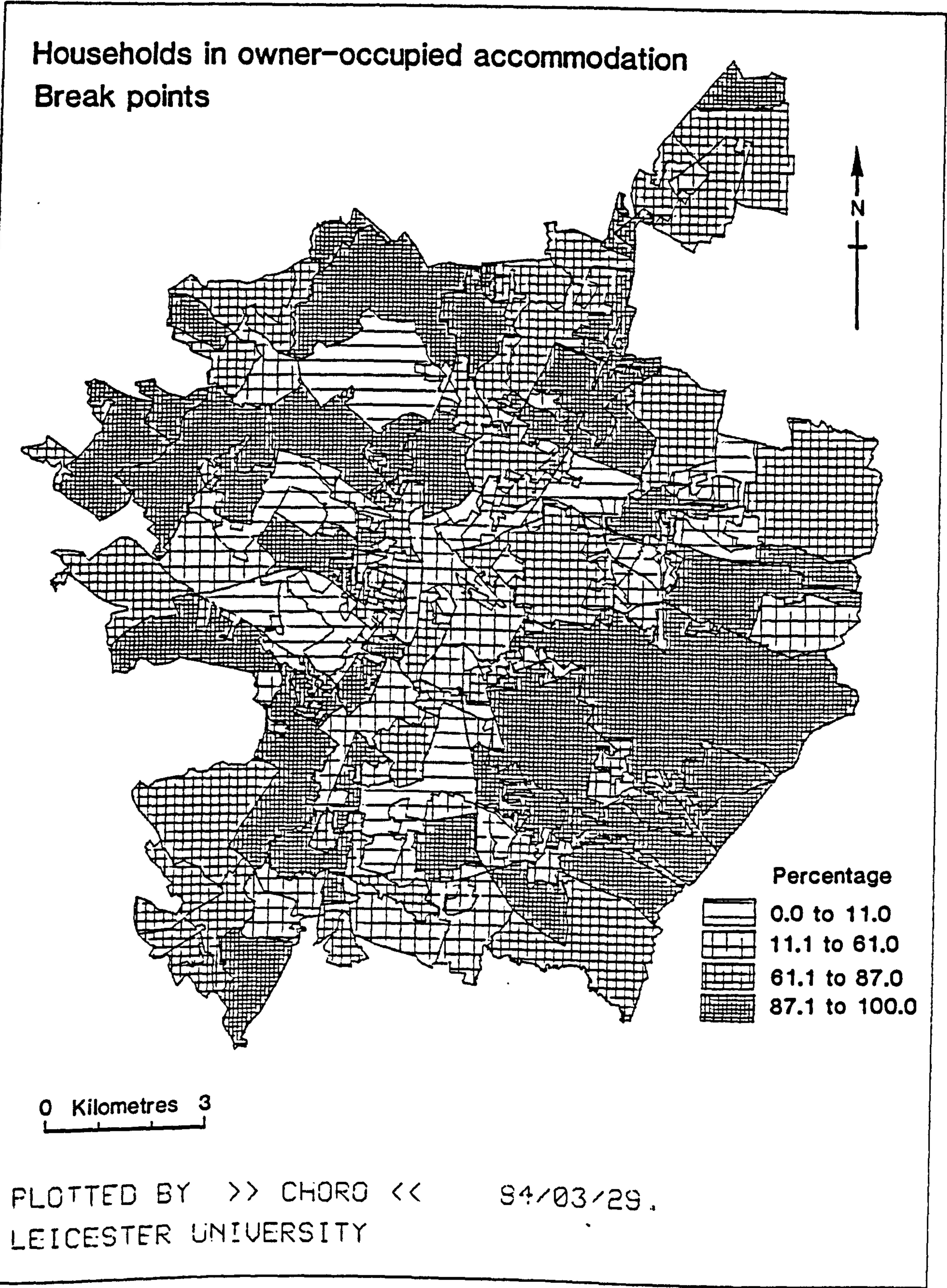
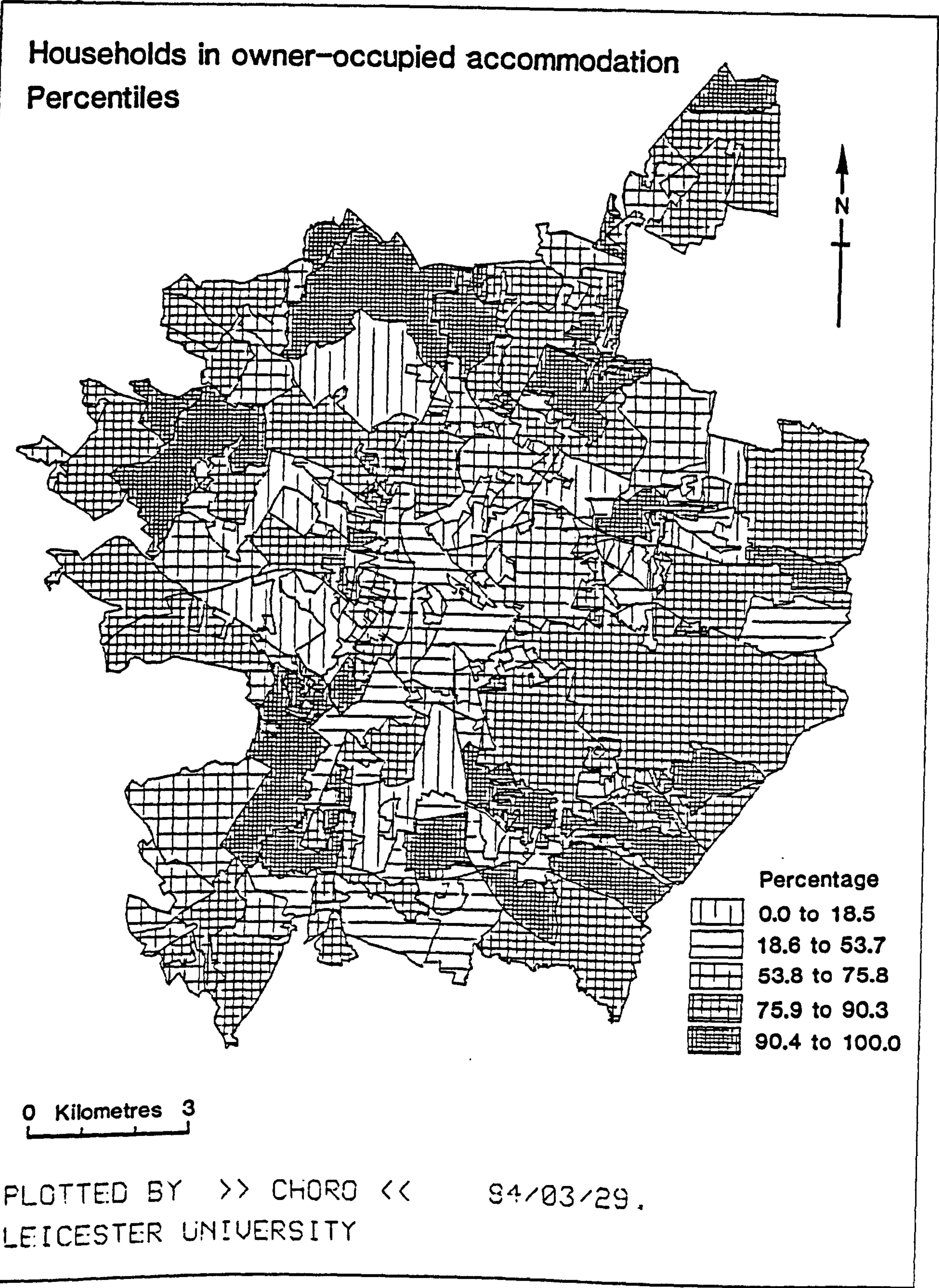


Figure A1.13



necessarily result in a loss of information but in fact enhances the readability of the map. Despite efforts to avoid it, the class limits chosen for the arithmetic progression (Figure A1.15) ensure that most data values lie in the lowest class. This is due to the extreme skewness of the data, and suggests that this method is particularly inappropriate. The percentiles and break points methods (Figures A1.16 and A1.17) have similar class limits and therefore produce comparable maps, both provide reasonable representations of the data. Neither the geometric progression, nor the mean and standard deviation systems (Figures A1.18 and A1.19) are suited to such highly skewed data.

It would have been preferable had one of the methods consistently performed better than the others so that it could be adopted for this thesis. Unfortunately this was not the case. On the whole, the 'percentiles', 'nested means' and 'break points' maps came out most favourably across all three variables, with the nested mean method having fewer significant disadvantages than the other two. All three of these methods accept the constraints of the closed ratio values, an important consideration when dealing with census data.

Table A1.2 summarises the main advantages and disadvantages of the different methods of class interval selection.

A1.4 Examining the effects of data transformation

The inability of the previous section to produce a 'best' result suggests that either the style of map or type of data value might be improved. An alternative to the choropleth map is the simple 'dot' map (used, for example, by Dixon (1979) in mapping population density in Portsmouth), or its more sophisticated relation, the proportional symbol map. The principle behind the latter is that map symbols should have visual importance equal to

Figure A1.14

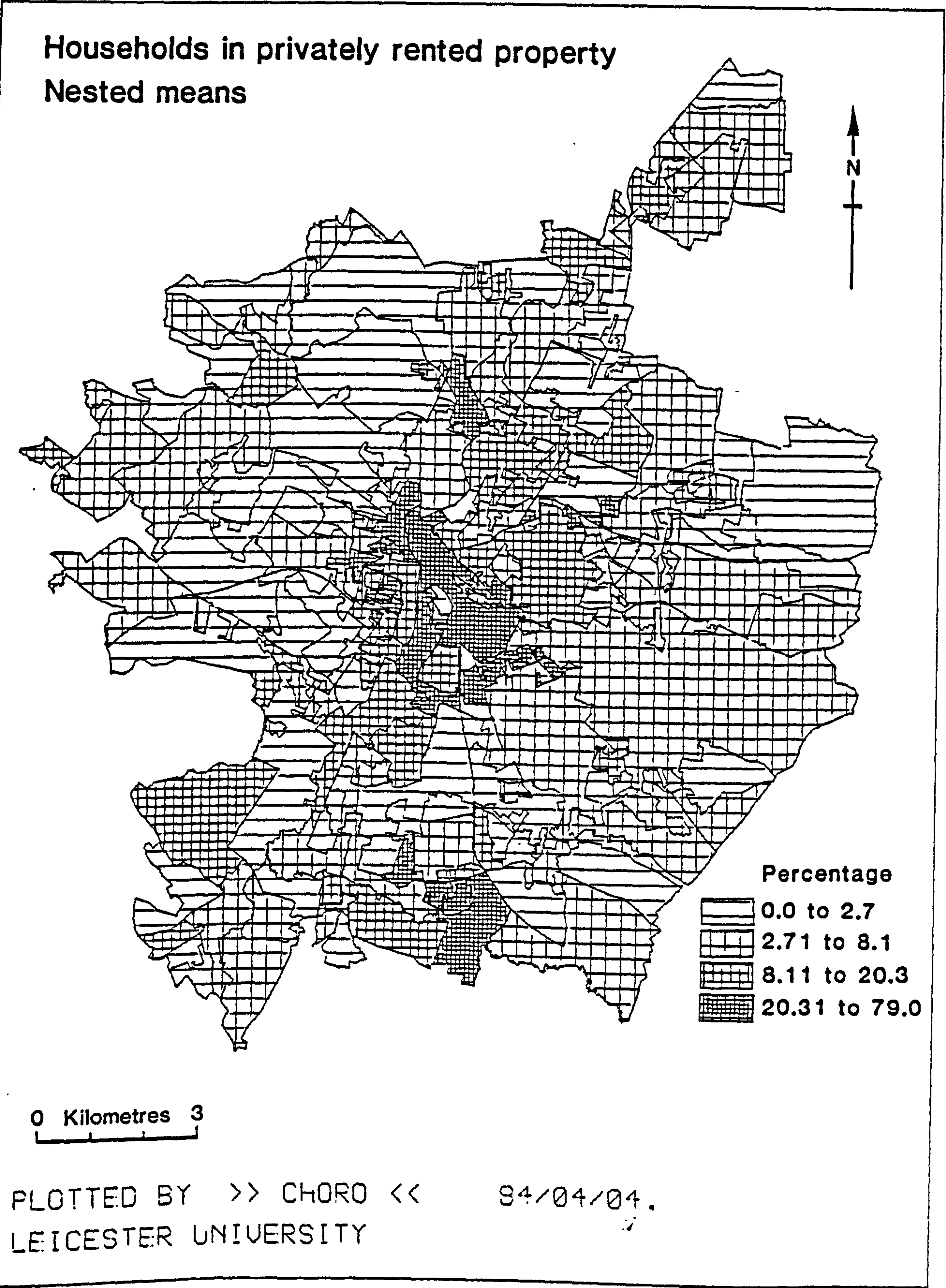


Figure A1.15

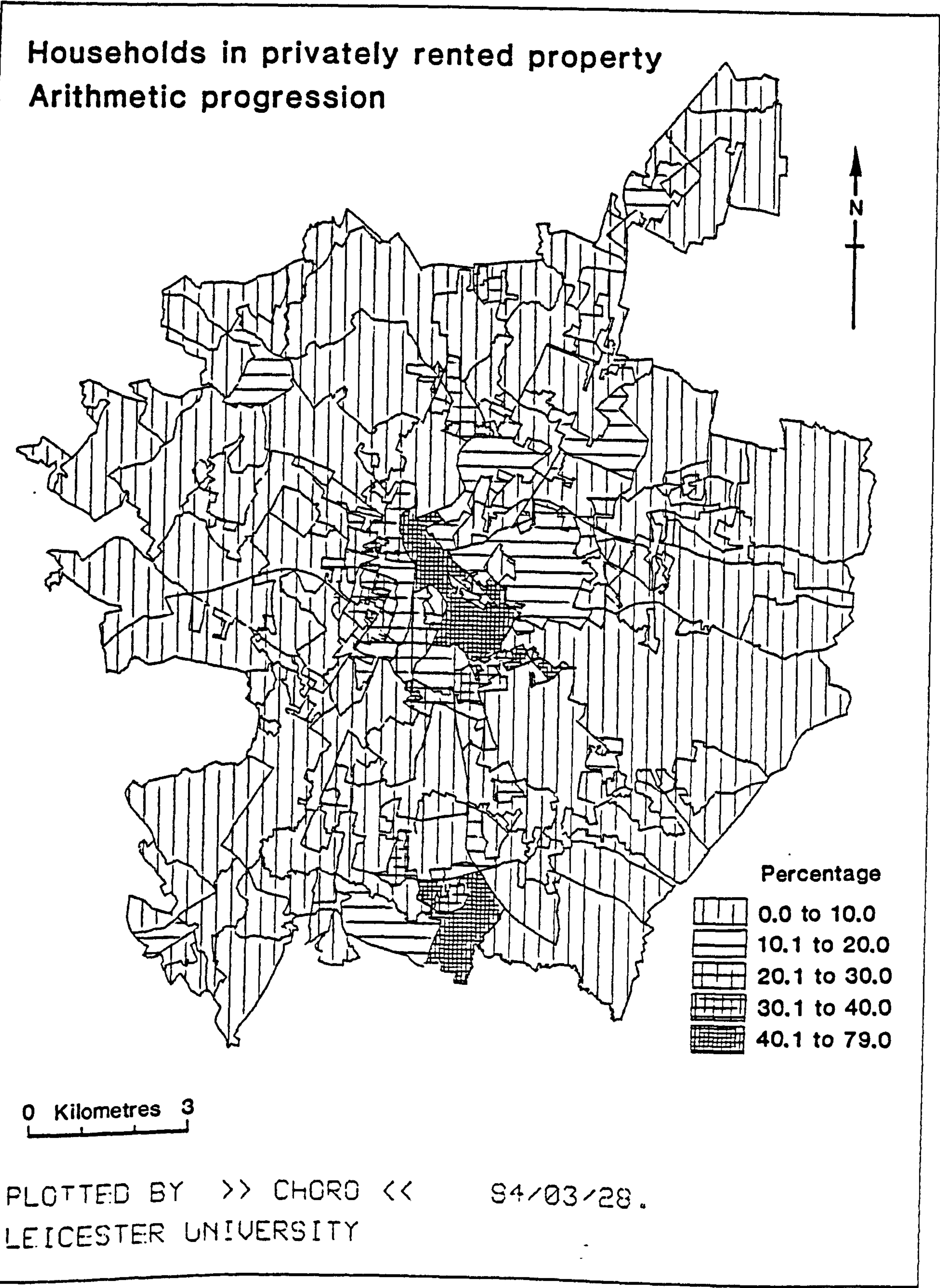
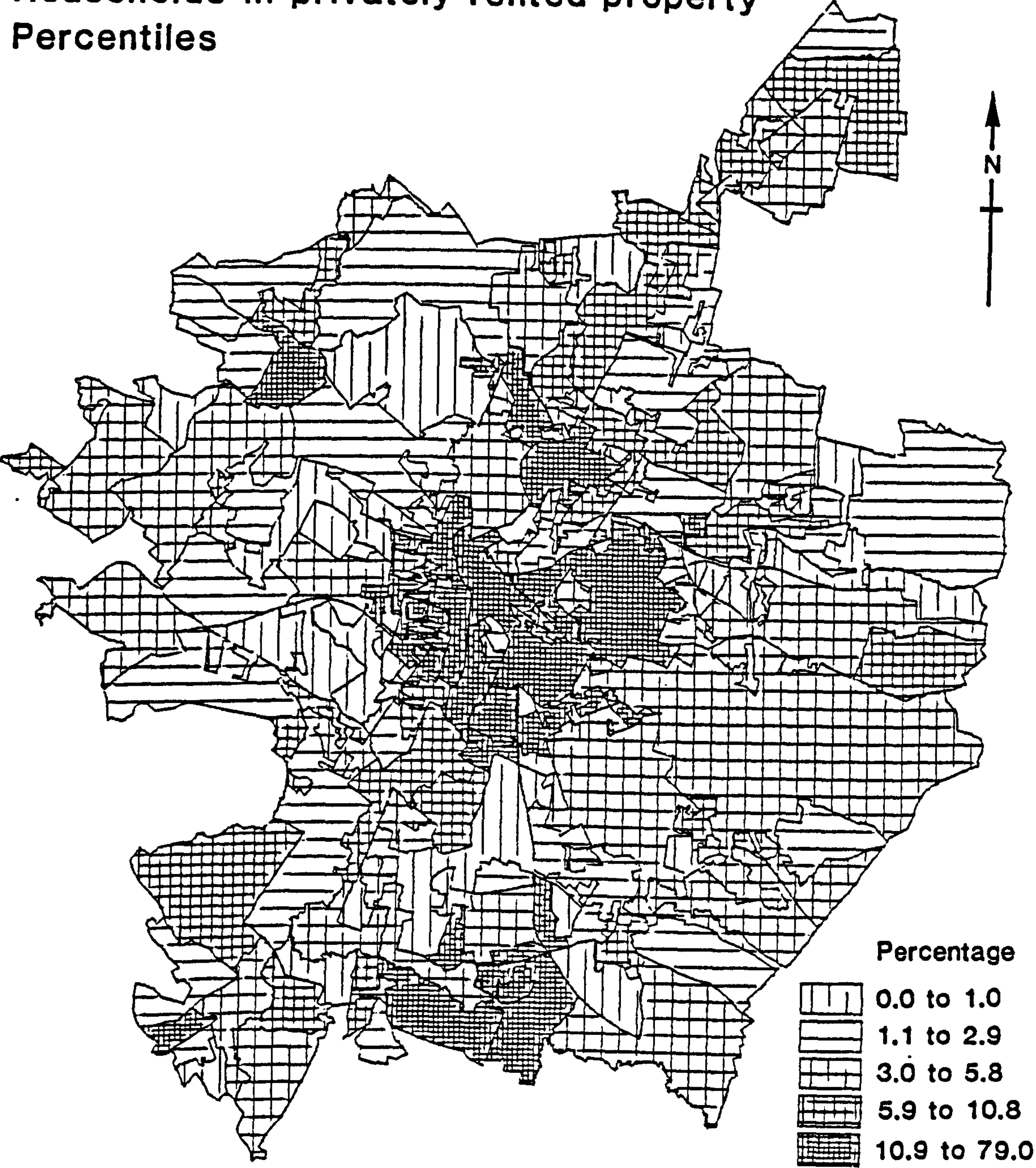


Figure A1.16

Households in privately rented property
Percentiles

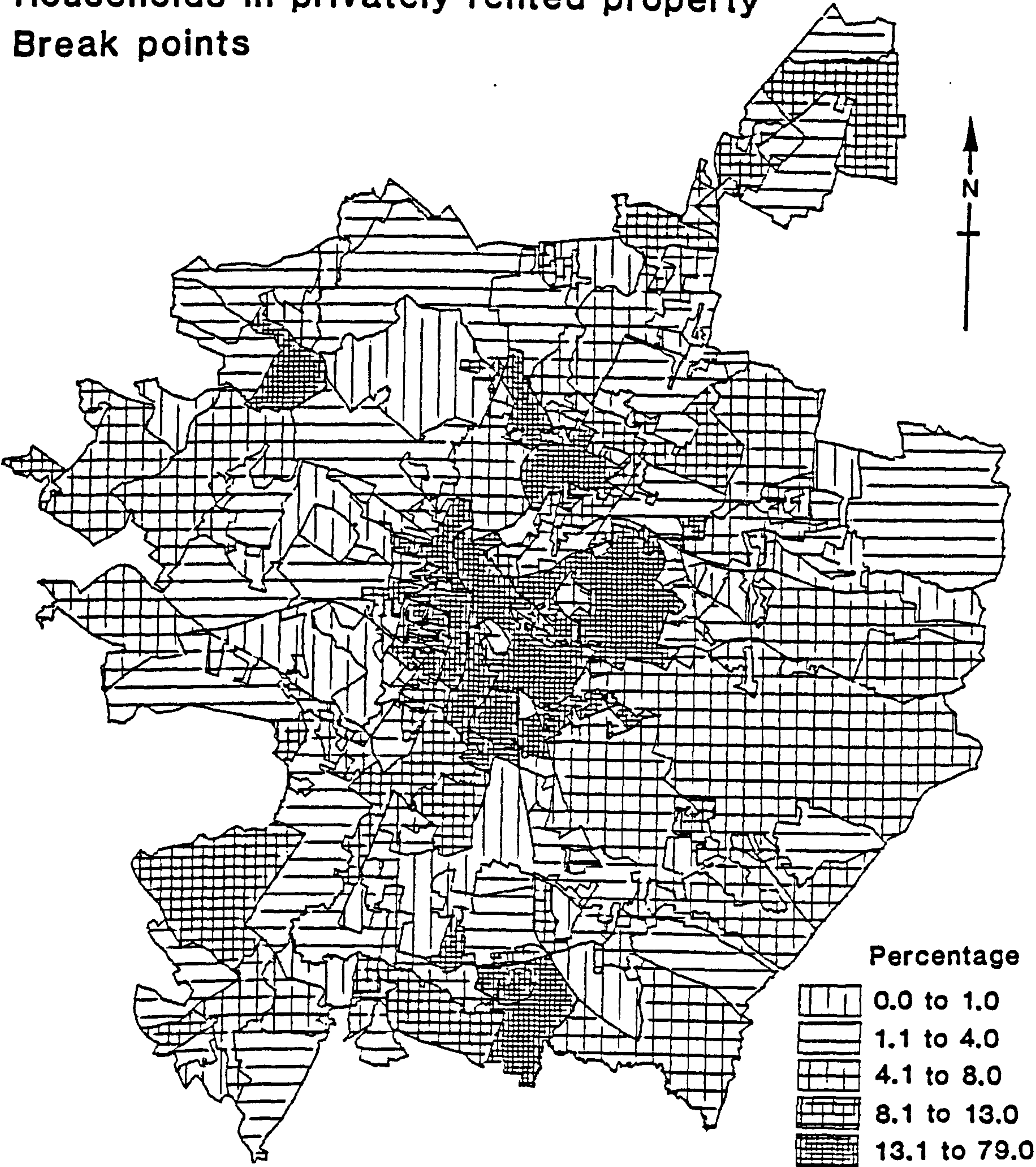


0 Kilometres 3

PLOTTED BY >> CHORO << 84/03/29.
LEICESTER UNIVERSITY

Figure A1.17

Households in privately rented property
Break points

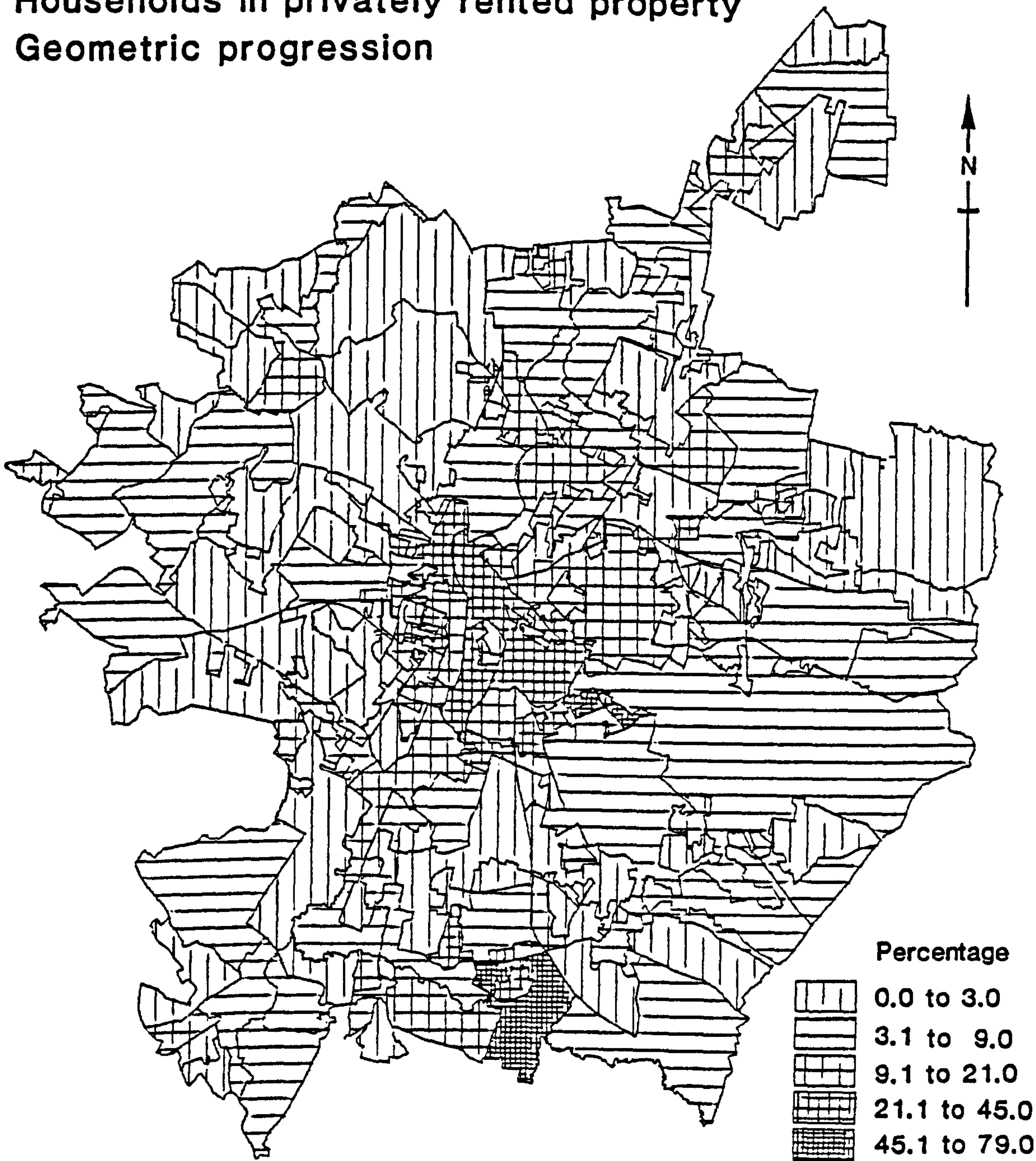


0 Kilometres 3

PLOTTED BY >> CHORO << 94/03/29.
LEICESTER UNIVERSITY

Figure A1.18

Households in privately rented property
Geometric progression

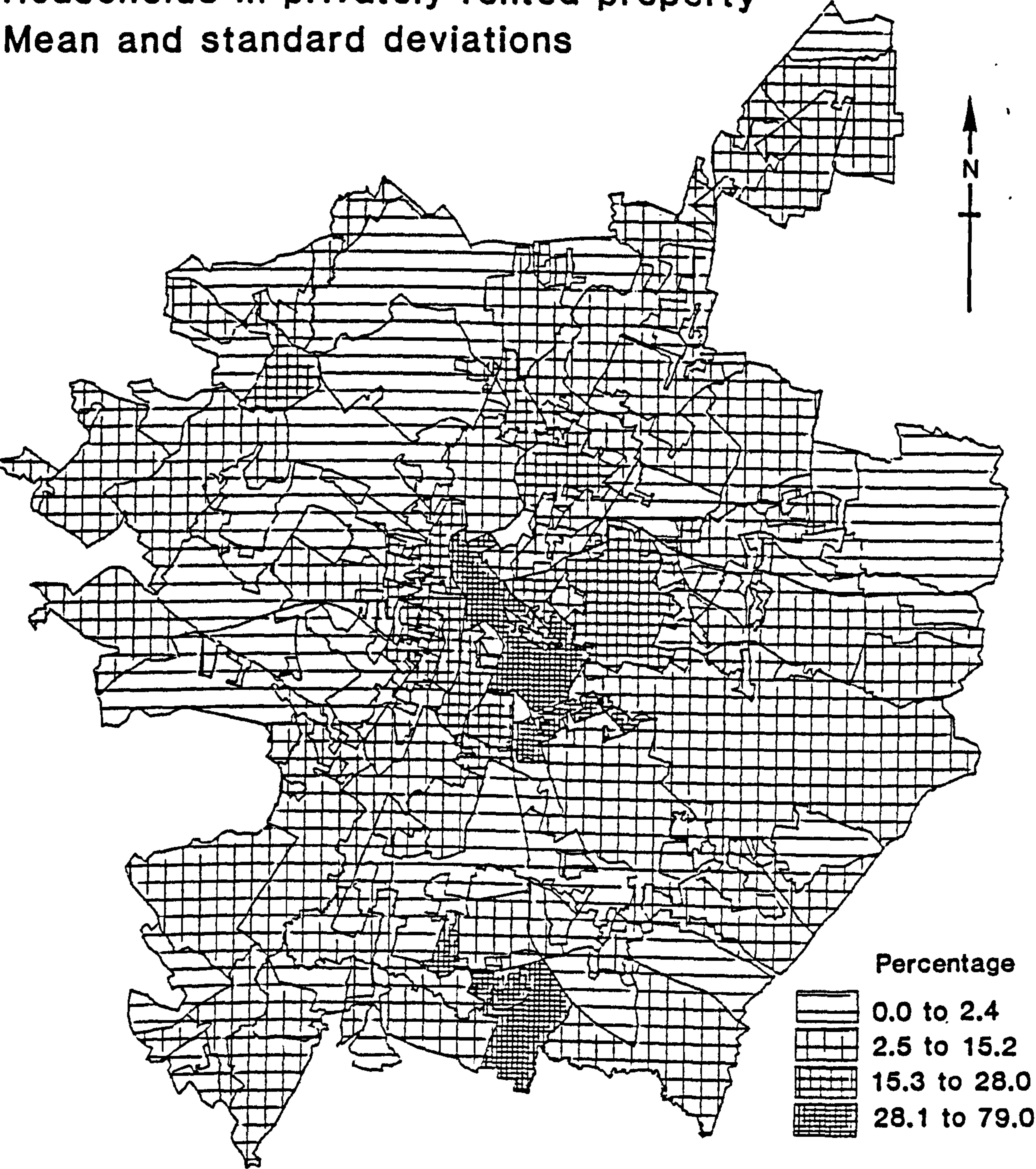


0 Kilometres 3

PLOTTED BY >> CHORO << 84/03/29.
LEICESTER UNIVERSITY

Figure A1.19

Households in privately rented property
Mean and standard deviations



0 Kilometres 3

PLOTTED BY >> CHORO << 84/03/29,
LEICESTER UNIVERSITY

Table A1.2 Assessment of methods of choosing class intervals.

1. BREAK POINTS (MULTIMODAL).

Advantages:

- direct response to the underlying data
- number of intervals prescribed by the data
- best suited to data which form 'natural' classes

Disadvantages:

- choice of interval is likely to be a subjective process unless the data form very clear clusters
- difficult to compare a series of variables

2. ARITHMETIC PROGRESSION

Advantages:

- conceptually straightforward, therefore easily understood and interpreted
- suited to data spread evenly across a range

Disadvantages:

- insensitive to the underlying data
- meaningless in any statistical sense

3. GEOMETRIC PROGRESSION

Advantages:

- may be used for standardising heavily skewed data (in theory)

Disadvantages:

- insensitive to the underlying data
- rapidly increasing size of class interval may cause problems at the upper end of the range

4. PERCENTILES

Advantages:

- even distribution across the classes produces a 'well-balanced' map
- enables identification of, say, the extreme 'x'% of areas for a given variable (easy to interpret)
- directly comparable across several variables

Disadvantages

- insensitive to the underlying data
- number of classes prescribed without reference to the statistical distribution

5. MEAN AND STANDARD DEVIATIONS

Advantages:

- based on well-tested statistical parameters
- identifies statistically significant extreme values
- comparable across variables of similar distribution

Disadvantages:

- unsuited to data not having a normal distribution
- class boundaries inhibited by the range of data values
- cannot calculate true standard deviations for ratio values

6. NESTED MEANS

Advantages:

- emphasizes extreme values
- direct response to the data
- comparable across variables of any distribution

Disadvantages:

- restricted to 2, 4, 8, 16... classes

the significance of the items represented (Jenks and Knos, 1961), where 'visual importance' is determined by the degree of contrast between symbols and is represented on the proportional symbol map by symbol size. In conventional mapping, proportional symbols are normally of graded sizes, each representing a range, or class, of data values. An advantage of computer cartography is that symbols may be directly proportional to the data values themselves and hence the need for identifying class boundaries is avoided. This facility enables the effects of other data transformations to be examined, such as the five listed below.

(i) Raw data (the 'absolute' value)

This is the number of observations in a spatial division sharing a given characteristic, for example, residents aged 0 to 4 in an enumeration district. Such a value is useful for example, to planners responsible for providing services, but it is critically dependant on both the size of the area and the total population within it.

(ii) Ratio of observations with a given characteristic to the total number of observations in an area ('population' ratio)

For example the proportion of people in an enumeration district that are aged 0 to 4. This is the value that has been used for all the analysis and mapping to date, it is the most commonly used method for standardising census variables.

(iii) Ratio of observations with a given characteristic to the size of the area in which they lie ('area' ratio).

An example of this is the number of owner-occupied households per square kilometre, hence it is a density. As with the previous value it is a means of standardisation, but one which is more effective where area size varies more than population size.

- (iv) **Ratio of observations with a given characteristic to both total observations and area size ('double density' index).**

The raw data are standardised for both population and area sizes. Thus, for a given absolute value, this index will be proportionally lower for either a large total population or a large area size. Put another way, to obtain a high value for the double density index, the ratio of the absolute value to population size must be large, and the area size small. In the context of census data this is clearly biased to those smaller EDs that are generally found near urban centres; it allocates far less importance to the large EDs which may be found in rural areas or within non-residential urban areas (such as parks or industrial estates). In cartographic terms, this index is appropriate for thematic mapping when one wishes to avoid giving large but low density areas undue visual prominence.

- (v) **The chi-square statistic.**

The signed chi-square measured for mapping was originally devised by Visvalingam (1976) for use with the U.K. 1971 census data. It has subsequently been the subject of much discussion and refinement (Visvalingam and Dewdney, 1977; Visvalingam, 1979; Jones and Kirby, 1980) and has been used in various census atlases (e.g. CRU/OPCS/GRO(s), 1980 and Maguire et al., 1983).

Essentially, the chi-square value is a measure of both the absolute and relative deviation of a characteristic within an area from the study area average. Thus, a positive chi-square value will result when a characteristic occurs more frequently in an area than would be expected given the global average, and a negative value when it

happens less frequently. The statistic incorporates a weighting such that for a given proportion with a characteristic, areas with large populations will have larger chi-square values than areas with small populations. The implicit assumption here is that while for the larger populations the numerical difference will of course be larger, unlike the ratio value, chi-square does not expect the magnitude of the difference to be a linear function of population size.

The chi-square value is calculated as follows:

$$\chi^2 = \frac{(O_1 - E_1)^2}{E_1} + \frac{(O_0 - E_0)^2}{E_0}$$

where:

O_1 = total number of observations with the given characteristic in an area (eg. number of people aged 0 to 4 in an ED).

E_1 = the number of observations that would be expected to have the characteristic if the area conformed to the global average (eg. total population of the ED multiplied by the overall proportion of 0 to 4 year olds in the study area).

O_0 and E_0 are equivalent values for those not sharing the characteristic (ie. people aged 5 or greater).

In order that the direction of the relationship between observed and expected values can be ascertained, a sign is taken from the difference between these values. Thus, if $O_1 > E_1$, then chi-square is positive, if $O_1 < E_1$ it is negative.

Tables A1.3 to A1.5 show key descriptive statistics for each of the census variables being investigated, Figures A1.20 to A1.22 show box plots and dispersion diagrams for the transformed

Table A1.3 Residents aged 45 to retirement – descriptive statistics

	Absolute value	Population ratio	Area ratio	Population/ area ratio	Chi-square value
Maximum	8508	0.439	9347.925	24.796	569.601
Minimum	3	0.019	15.556	0.012	-472.379
Median	147	0.194	814.332	0.886	0.0
Mean	363.048	0.196	916.633	1.552	0.807
Standard deviation	775.827	0.051	811.221	2.304	58.924

Table A1.4 Households in owner-occupied accommodation – descriptive statistics

	Absolute value	Population ratio	Area ratio	Population/ area ratio	Chi-square value
Maximum	11174	1.0	4312.617	45.257	4183.53
Minimum	0	0	0	0	-7071.234
Median	159	0.64	694.525	1.378	2.154
Mean	408.427	0.575	973.469	4.044	-98.928
Standard deviation	982.274	0.331	927.01	5.851	885.790

Table A1.5 Households in privately rented property – descriptive statistics

	Absolute value	Population ratio	Area ratio	Population/ area ratio	Chi-square value
Maximum	1738	0.790	3745.744	22.272	3330.717
Minimum	0	0	0	0	-476.455
Median	13	0.043	42.694	0.129	-5.268
Mean	53.943	0.088	227.52	1.028	61.461
Standard deviation	155.46	0.128	523.026	3.056	348.19

Figure A1.20 Residents aged 45 to retirement Dispersion Diagrams

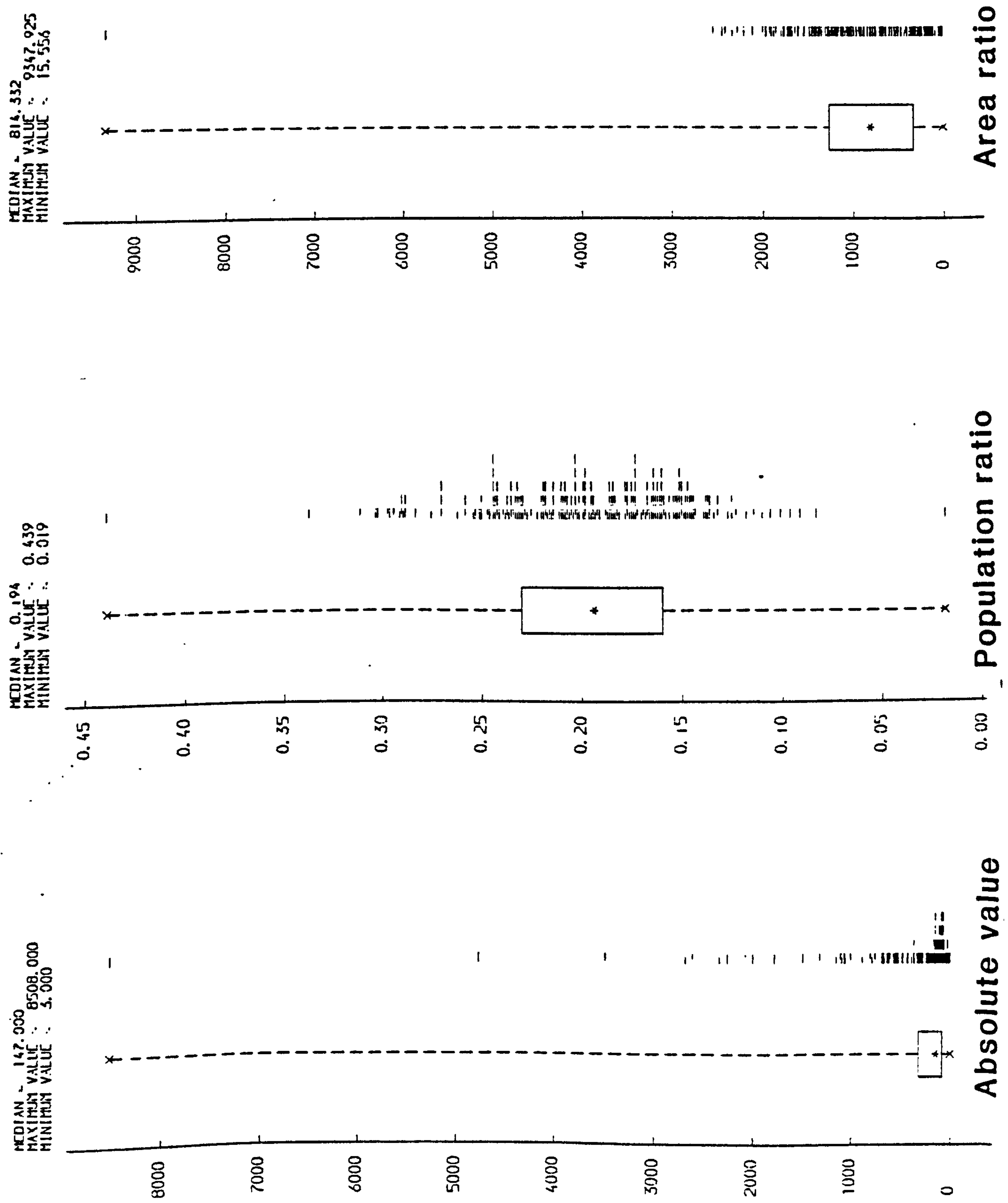


Figure A1.20 (continued)

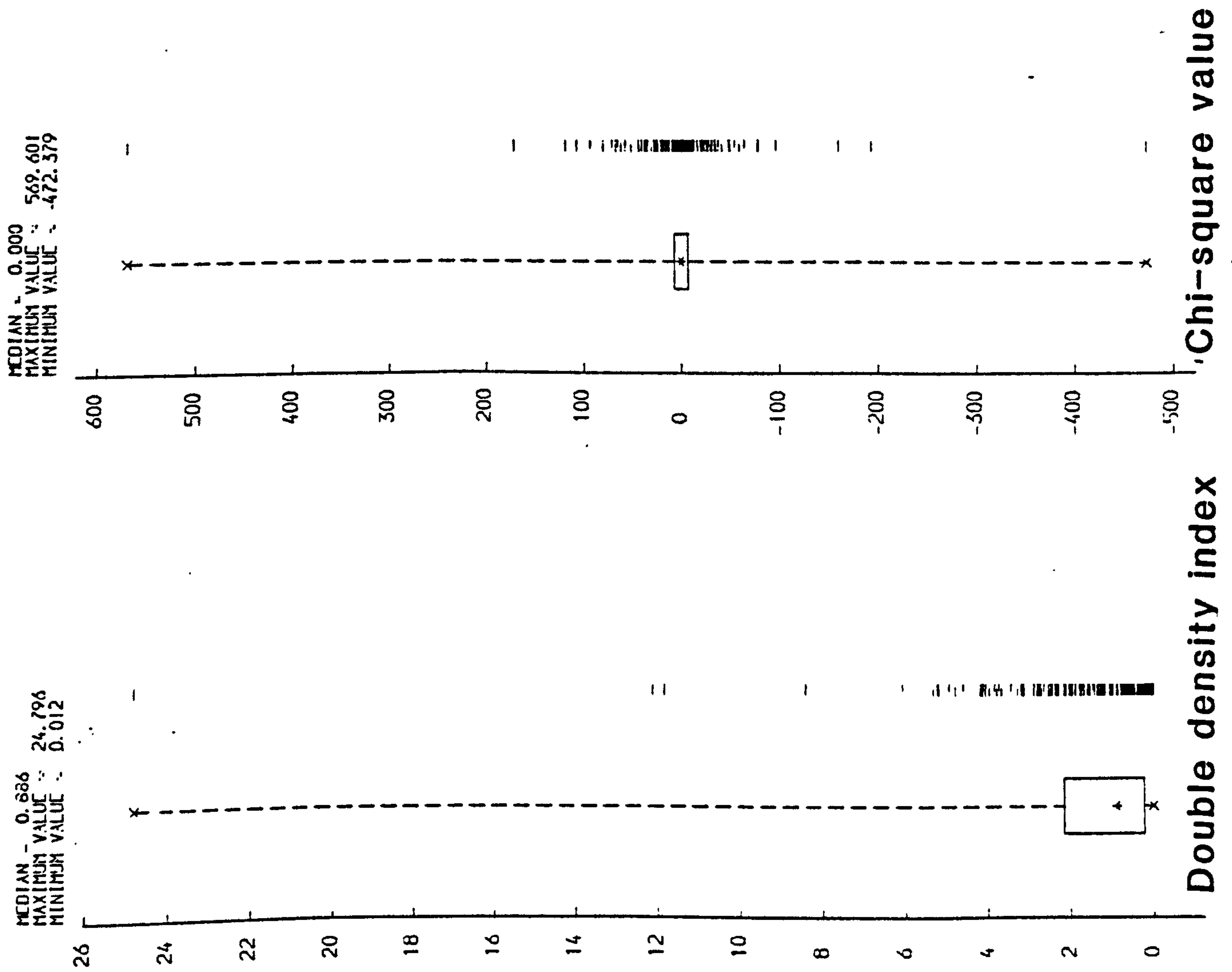


Figure A1.21 Households in owner-occupied accommodation Dispersion diagrams

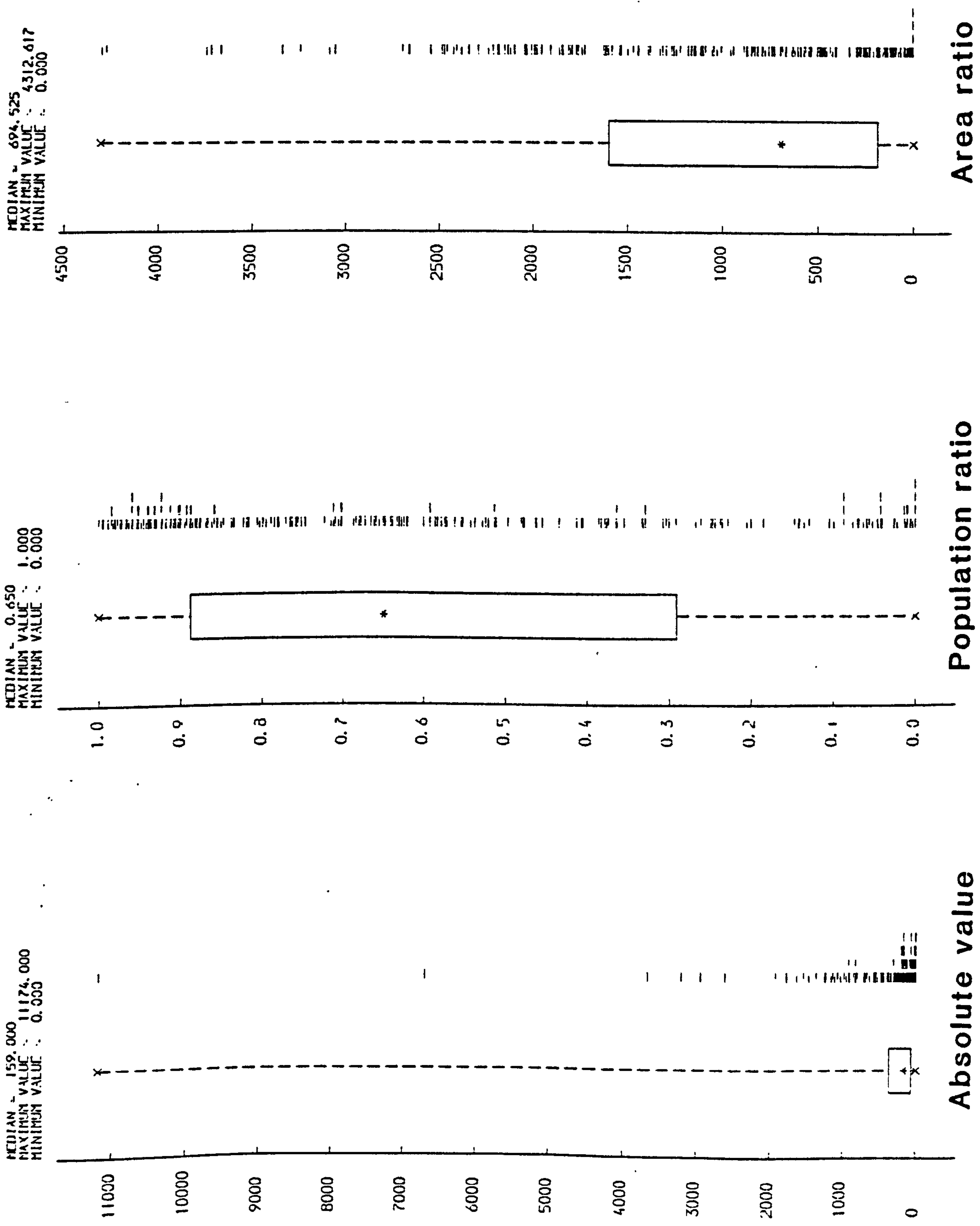


Figure A1.21 (continued)

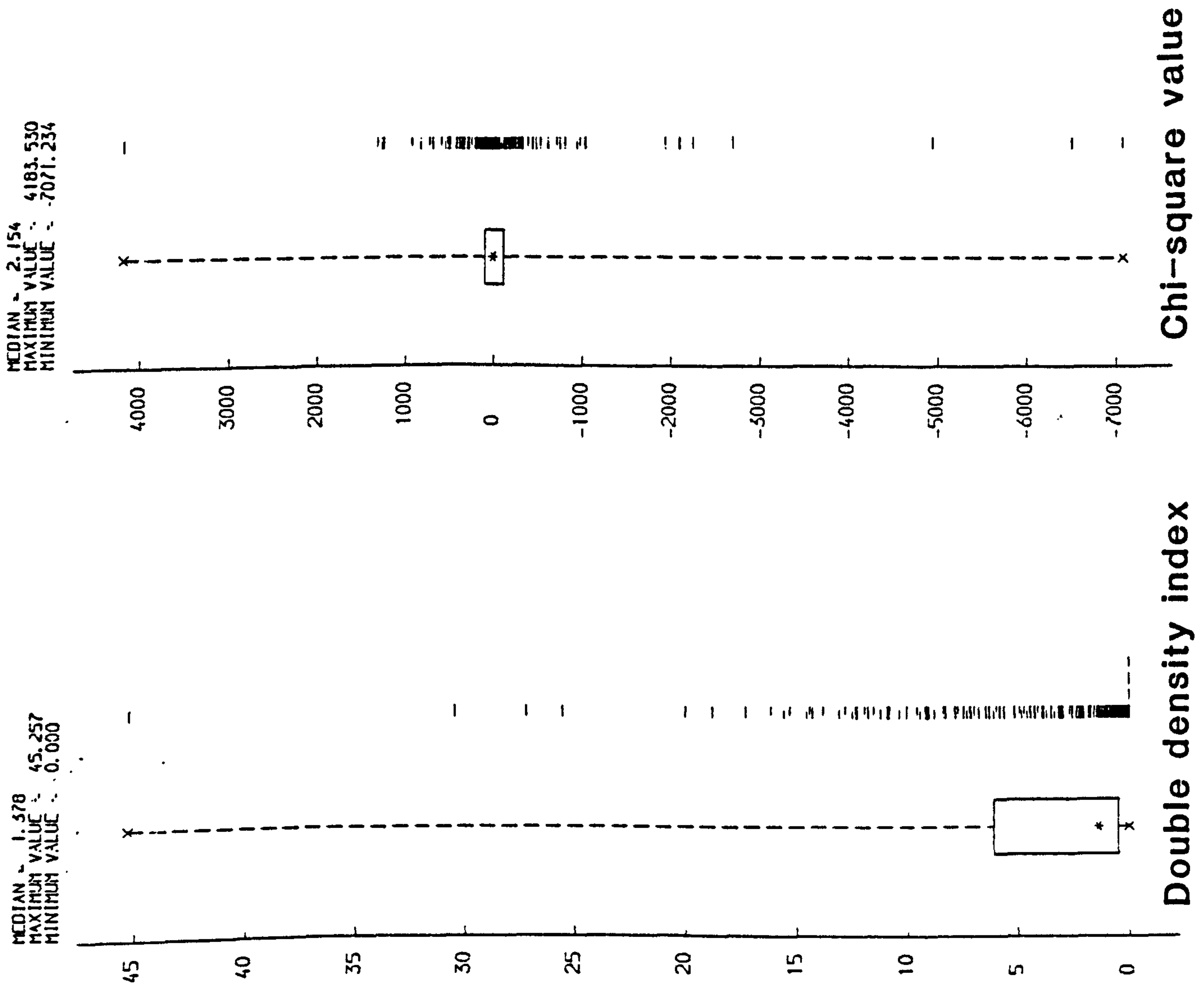


Figure A1.22 Households in privately rented accommodation Dispersion diagrams

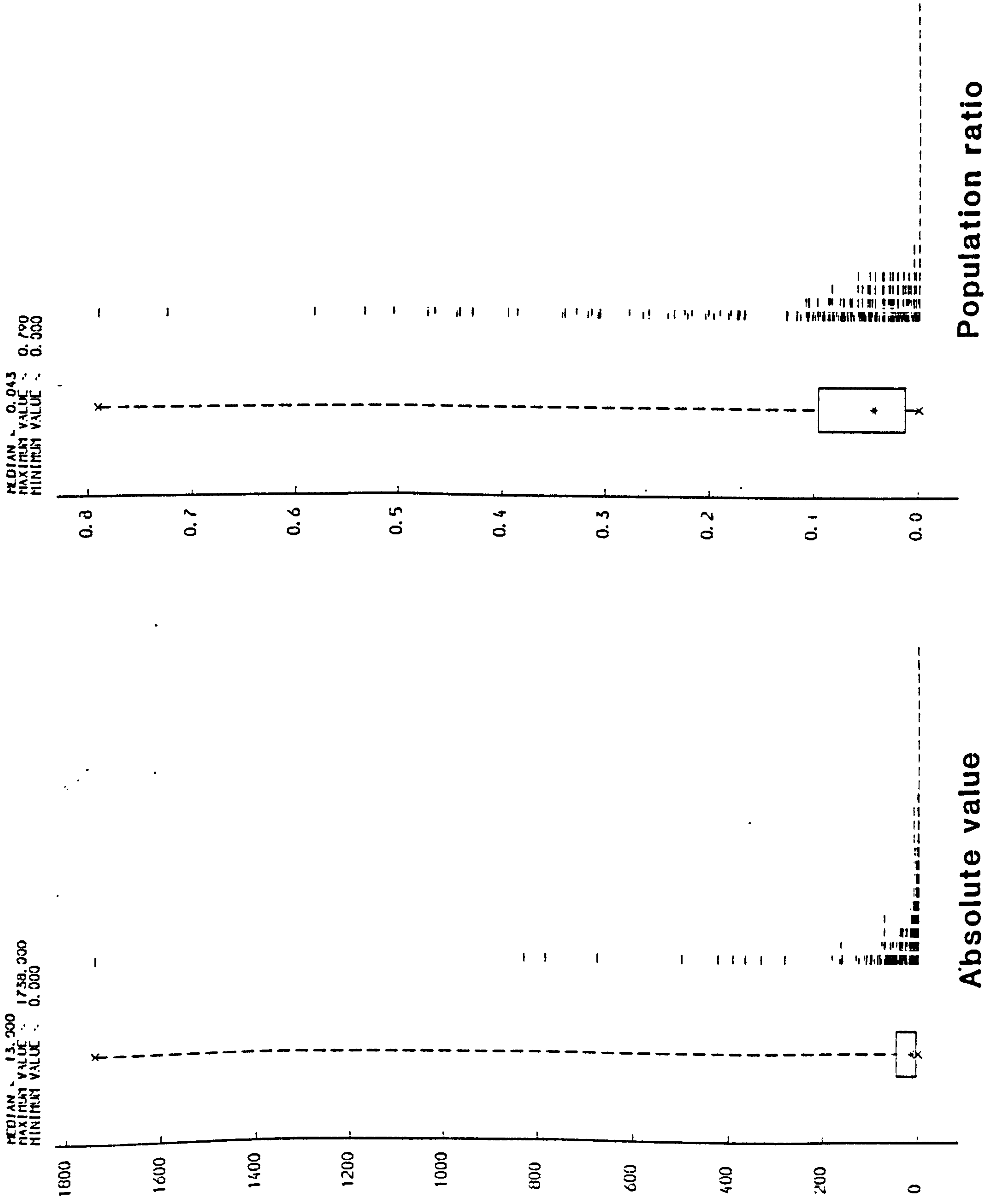


Figure A1.22 (continued)

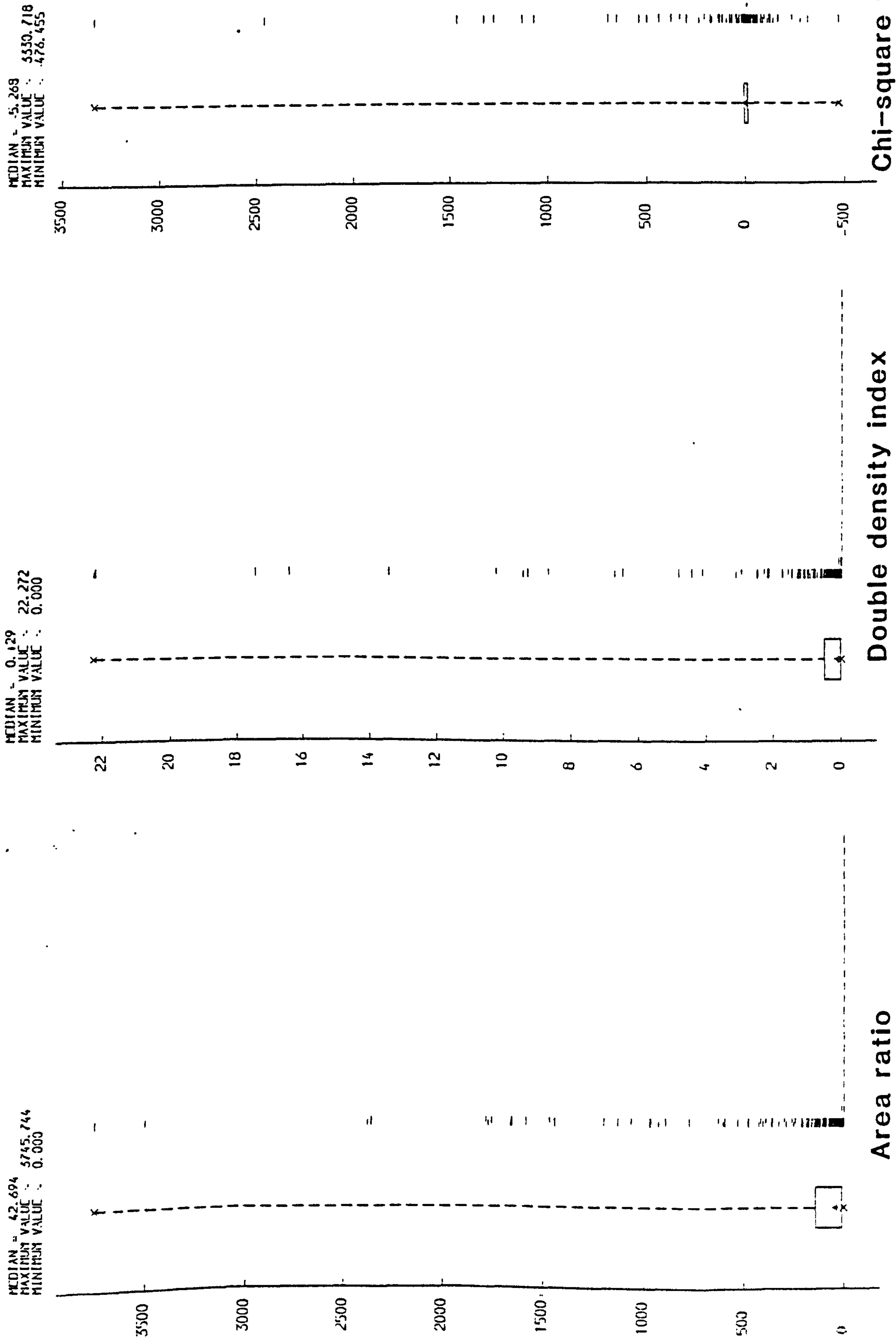


Figure A1.23 Residents aged 45 to retirement
Histograms of transformed data values

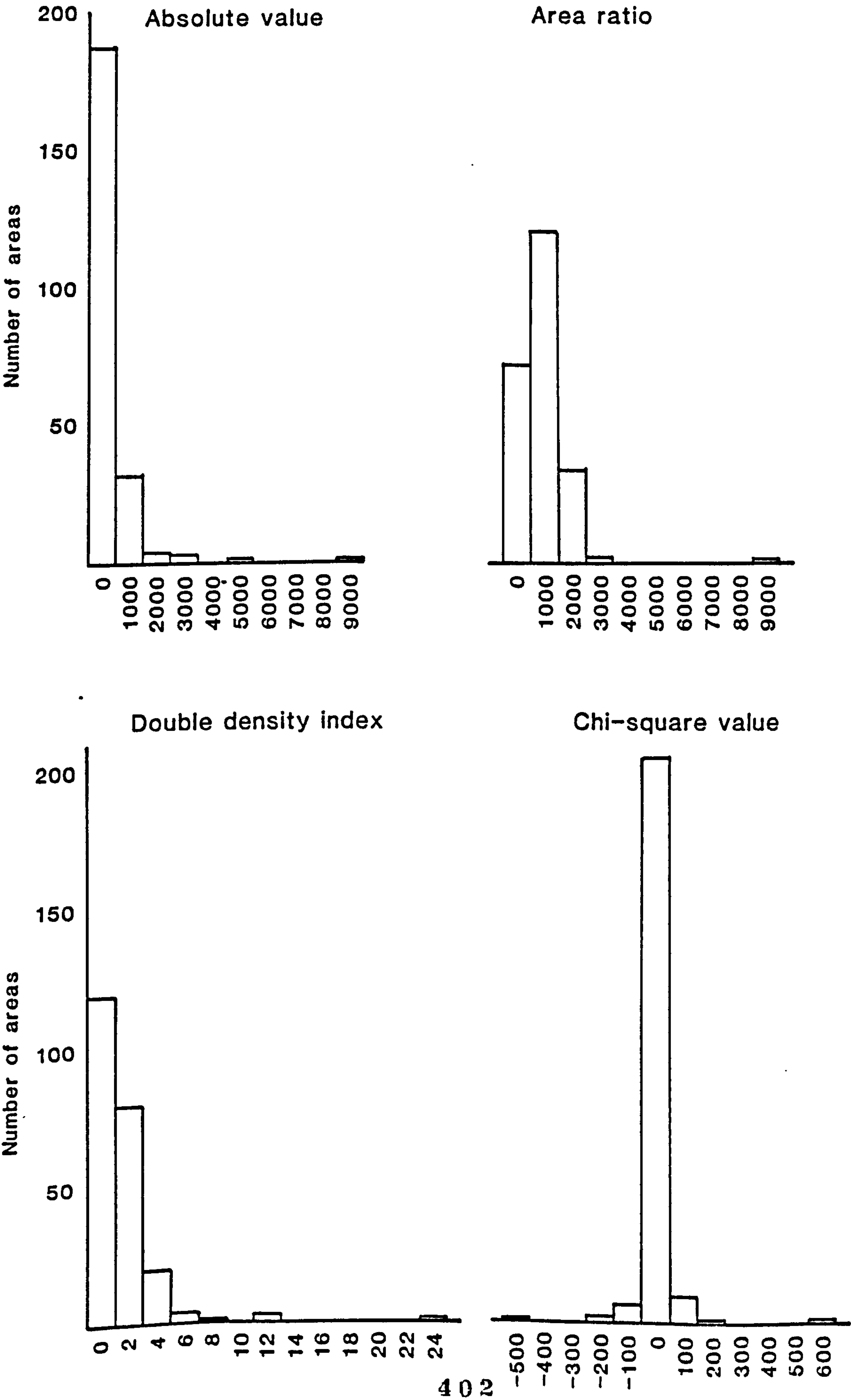


Figure A1.24
Households in owner-occupied accommodation
Histograms of transformed data values

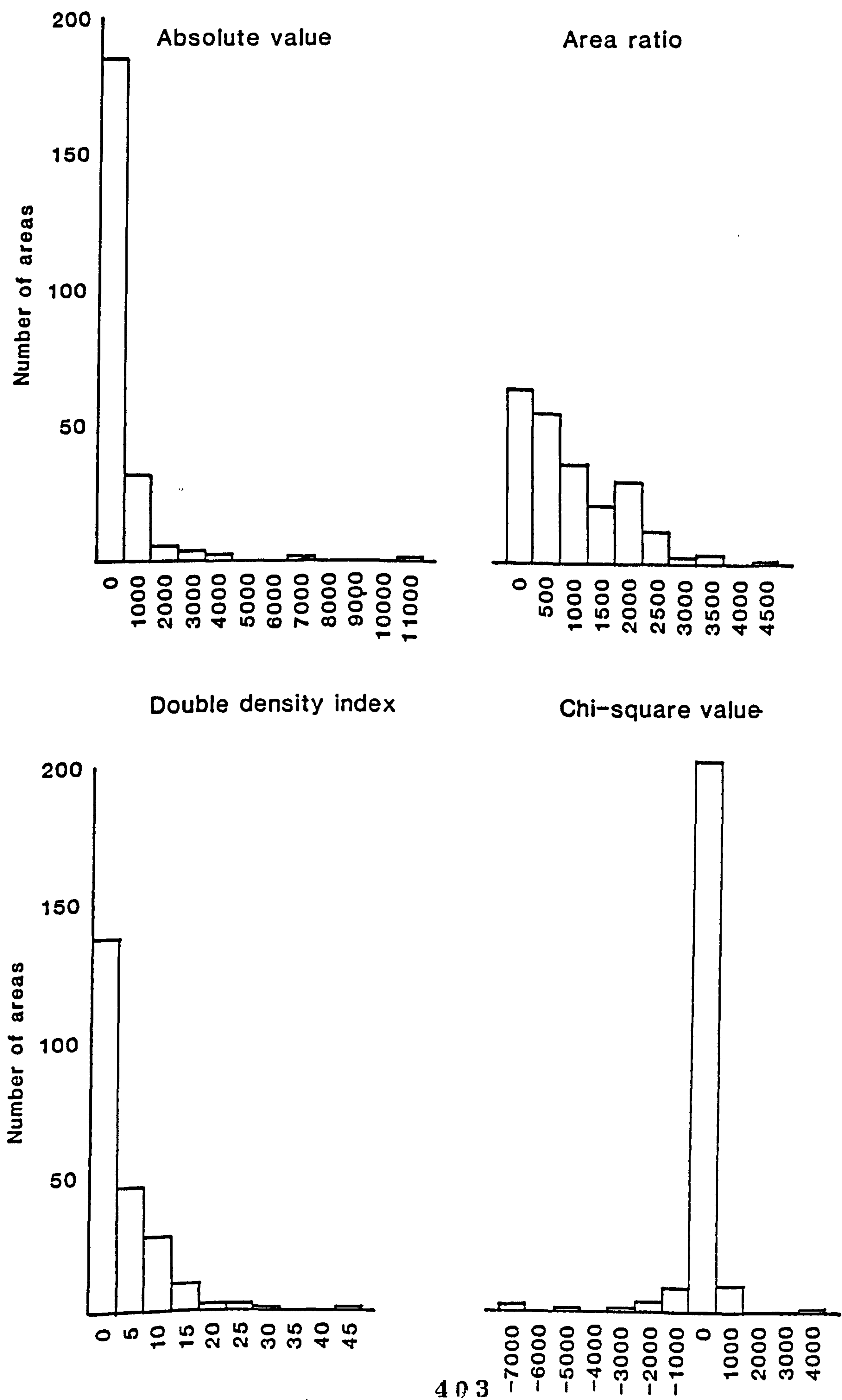
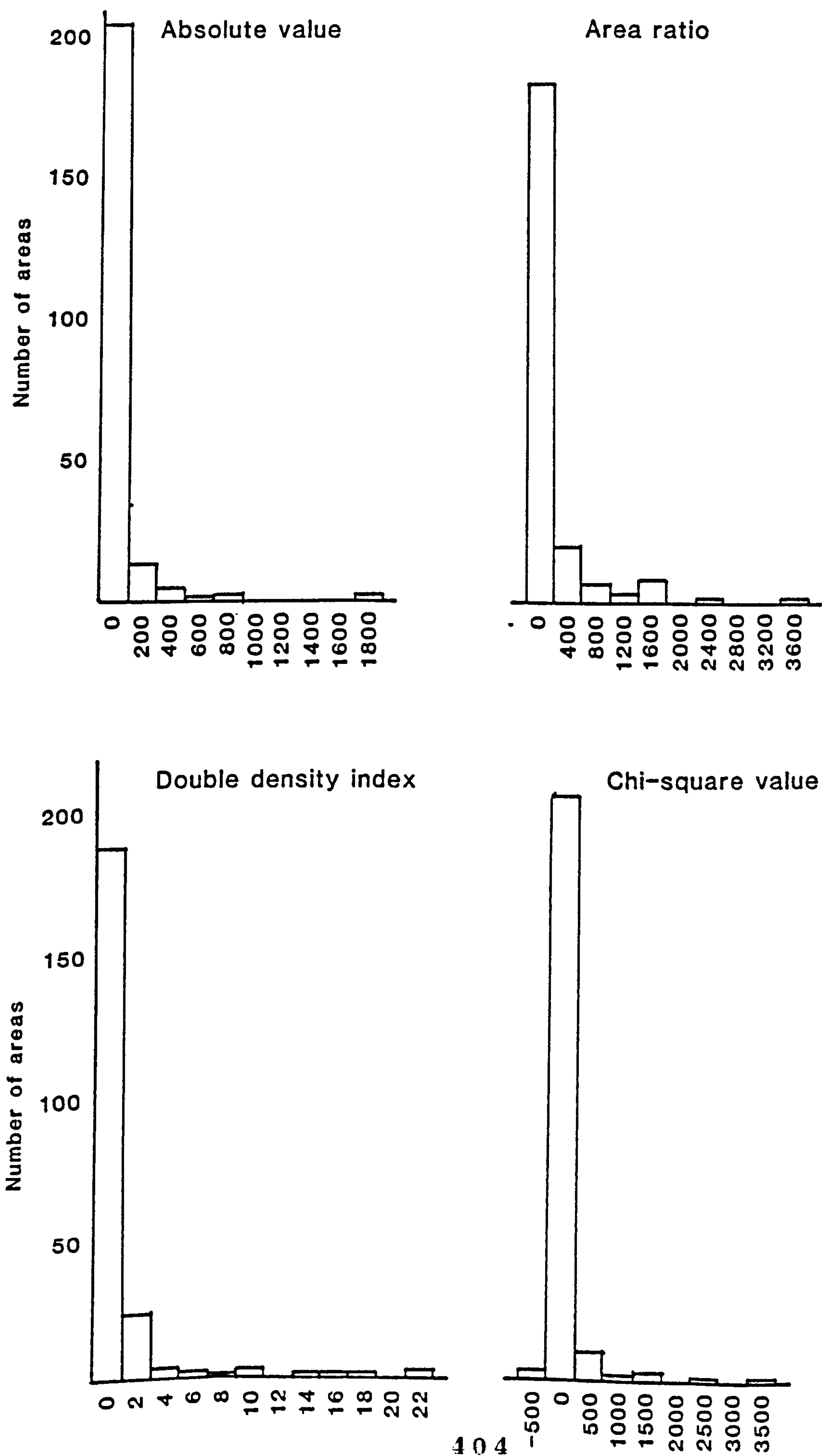


Figure A1.25 Households in privately rented property
Histograms of transformed data values



Before plotting proportional symbols for each of these variables it was necessary to derive individual point locations to represent the new small areas. If the areas mapped had been the original EDs it would have been possible to extract centroid values directly from the census data via SASPAC. However, having combined many EDs in the clustering process, this was no longer possible. Instead a program was written to calculate the mathematical centroid of each area based on a string of coordinates describing its boundary. At the same time the area (in square kilometres) of each new zone was calculated.

The remaining decisions concerned the shape and size of the proportional symbols. While the literature on both of these problems is extensive, practical considerations determined the choice of symbols here. In the interests of plotting time the shape of the symbols had to be simple and in practice was restricted to standard characters from the graphics font available in GHOST80. Symbol size was ultimately determined by the data, within limits arbitrarily defined to look reasonable on the map. Thus symbol sizes ranged from a minimum scaled height of 0.15km to a maximum of 1.65km. The following formula was used:

$$\text{Height of symbol} = 0.15 + \left(0.15 \times \left(\left(\frac{100}{V_{\max} - V_{\min}} \right) \times (V - V_{\min}) \right) \right)$$

where V_{\max} = maximum data value

V_{\min} = minimum data value

V = current data value

In this way the data values were scaled firstly to values within a range of 0 to 100 and secondly to symbol plotting sizes between the limits set above. Note that the height rather than the area of the symbol was taken as proportional to the data value, in addition to being computationally simpler, it was felt that the different shapes could more readily be compared in this manner.

Figures A1.26 to A1.28 show the population ratio values for the variable 'Households in owner-occupied accommodation'; squares, triangles and octagons are plotted. On each of the maps there are areas in which symbols overlap considerably, making individual values difficult to distinguish. This is inevitable with this type of mapping and since in itself it conveys the correct impression of high density it is not considered to be a major problem. The triangle was judged to minimise the confusion on the map and was therefore used for the remainder of the plotting.

A1.5 Comparing maps of transformed data values.

A total of fifteen maps were plotted, showing the five transformations of each of the three variables. These have not been included since as described below they were not particularly informative. The objectives of the exercise were two-fold: firstly to examine the effects of data transformation on maps of census data; and secondly, to assess the validity of the proportional symbol map for displaying census variables.

Two general observations may be made. First, the statistical distribution of the underlying data is the most important determinant of the eventual mapped pattern, therefore previous comments also apply here. Second, it is self-evident that clustering in the statistical distribution produces symbols of similar sizes and the direction of any skewness greatly affects the readability of the map. A negatively skewed distribution produces many small symbols which may be easily distinguished, a centrally clustered distribution enables extremely high and low data values to be identified, but a positively skewed distribution produces mainly large and therefore overlapping symbols which result in a confused picture.

Of the three variables, 'Households in owner-occupied

Figure A1.26

Comparing census plotting symbols for variables

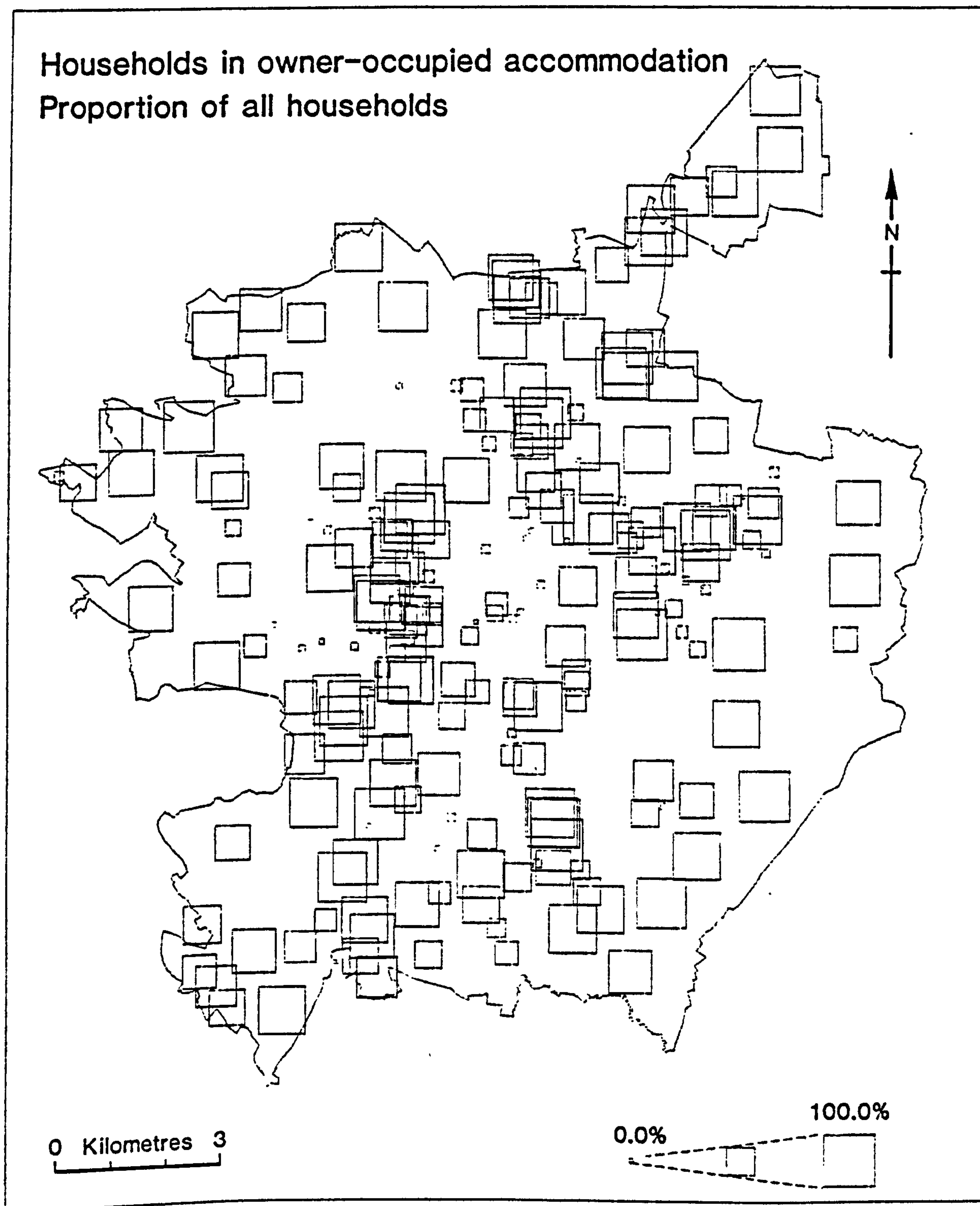


Figure A1.27

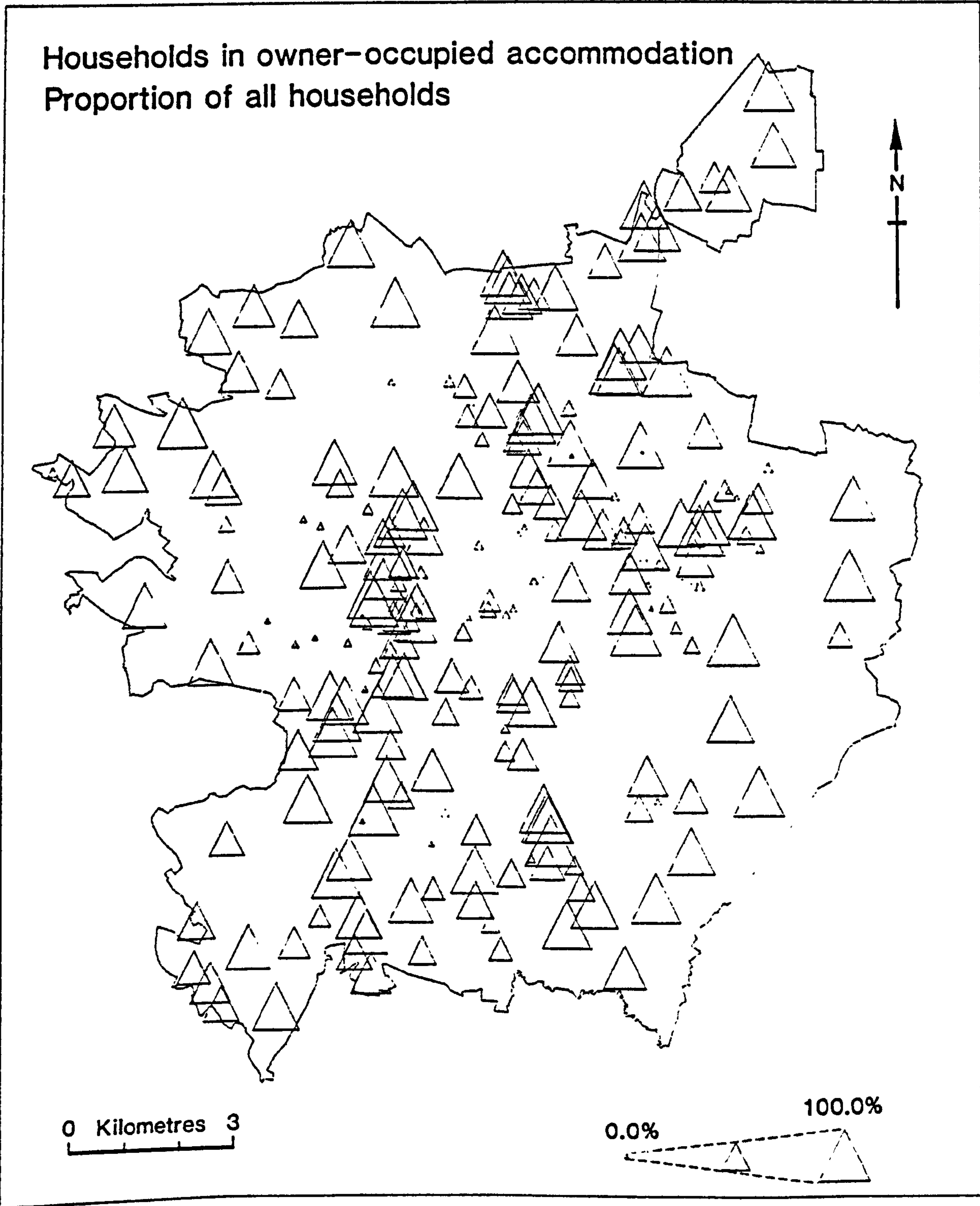
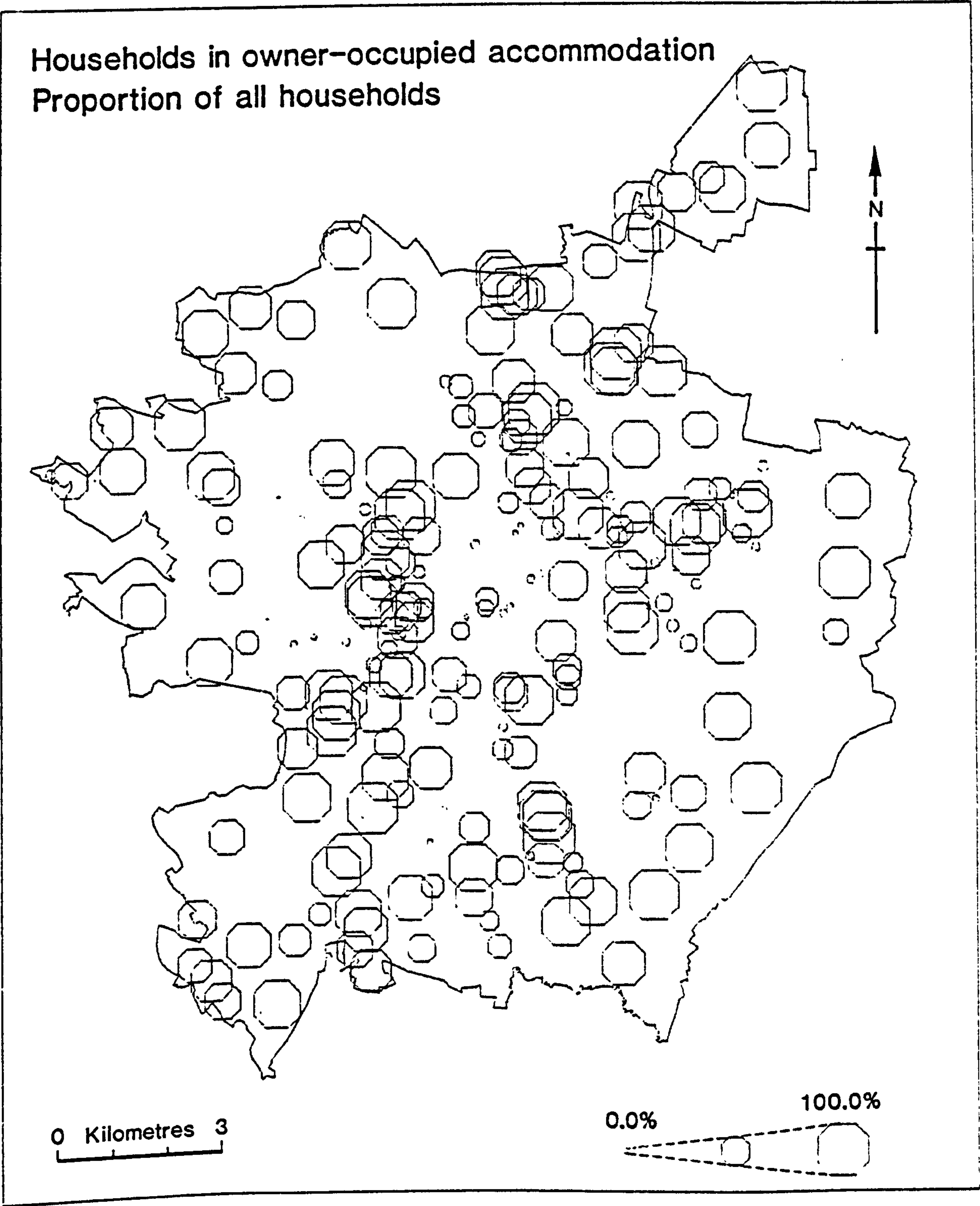


Figure A1.28



accommodation' had the greatest variation in its distributions when transformed, it therefore showed the greatest differences when mapped. Among the transformations, absolute values for the three variables exhibited similar mapped patterns. This was due to the overall spatial distribution of the study area's population and emphasises the unhelpfulness of this transformation. Both the area ratio and to a lesser extent the population ratio resembled the double density index in terms of its mapped pattern. This is reasonable given the derivation of the latter value.

Interpretation of the maps is easiest when most values lie to the middle of the range and extreme values are clearly represented. Care must be taken however, to give equal significance to both high and low values despite the visual dominance of the larger symbols. The presence of large areas containing no symbols at all should also be viewed with caution; this is due to the non-uniform spatial distribution of the centroid points. The relative size of the symbols is of greater importance than their juxtaposition.

In conclusion, the usefulness of any transformation probably depends mostly on the purpose of the map. Thus the area ratio is appropriate for studying areal concentration, while the population ratio is suitable for examining the distribution of a characteristic within a global population. The double density index is conceptually more complex and is therefore difficult to evaluate and hence unhelpful. The major attraction of the chi-square value for mapping is the facility of comparing values with an objectively derived 'expected' value. That this cannot satisfactorily be shown by the proportional symbol method is disappointing and suggests that an alternative method of display should be found.

Because of this the use of continuously graded proportional symbols is therefore considered inappropriate to census data. Stepped symbol sizes might have enhanced the legibility of the maps but they would have entailed selecting class limits for each size and defeated the purpose of the exercise. Variable transformation produces interesting results although three of the values, the absolute value, the areal density and the double density index, were not considered helpful. The population ratio is the conventional value for processing census data, and since it has been used already for the multivariate analyses it would be suitable for mapping. Alternatively, the chi-square value seems to offer the advantages of the population ratio on which it is based without the disadvantages for unevenly sized spatial units. It is therefore suggested that a final experiment should be conducted - a choropleth type display of chi-square data values.

A1.6 A recent development in mapping census data - the chi-square statistic.

The calculation of the chi-square measure has already been described, in this section each of the three census variables is mapped in the choropleth style and assessed according to the criteria previously stated (see section A1.3).

The use of the chi-square measure in this project enables the testing of Evans' (1977) 'exogenous' approach to choosing class intervals, in which class limits are chosen at meaningful threshold values. For the chi-square statistic meaningful values are found at levels of significance in the chi-square distribution. For 1 degree of freedom and critical thresholds of 10% and 0.1% the resultant class limits are 2.71 and 10.83. Since a signed measure is used for mapping the full range of classes becomes:

- Class 1 -10.83 or less
- 2 -10.82 to -2.71
- 3 -2.7 to 2.71
- 4 2.72 to 10.83
- 5 Greater than 10.83

These intervals are, of course, the same for all three variables.

Figures A1.29 to A1.31 can be compared with the earlier choropleth maps of section A1.3. When interpreting these the following points should be remembered since conventional ratio values are not being used. First, the middle category contains areas with a proportion of the characteristic that is not significantly different from that of the study area as a whole, given the population size of those areas. These areas may therefore be said to contain an 'expected' proportion. The outermost categories, on the other hand, contain areas with significantly more or less of the characteristic, given their size, than the global average. Hence they may be said to have much more or less than expected. The 'mean value' shown on each map represents the overall proportion of the study area population having the characteristic. It is shown to indicate the approximate magnitude of the 'expected' value.

On all three maps the chi-square statistic tends to span all five categories, having far more areas in the lowest and highest classes than would be normal for a ratio measure. In Figure A1.29, above average proportions of the older working age group can be seen in the middle to outer suburbs and a low proportion 'wedge' runs from south west to northeast. Compared to the earlier ratio-based maps, the chi-square map is certainly more useful than the arithmetic or geometric progressions, the 'break points' map or the mean and standard deviations map. It compares favourably with the percentiles and nested means maps both for

Figure A1.29

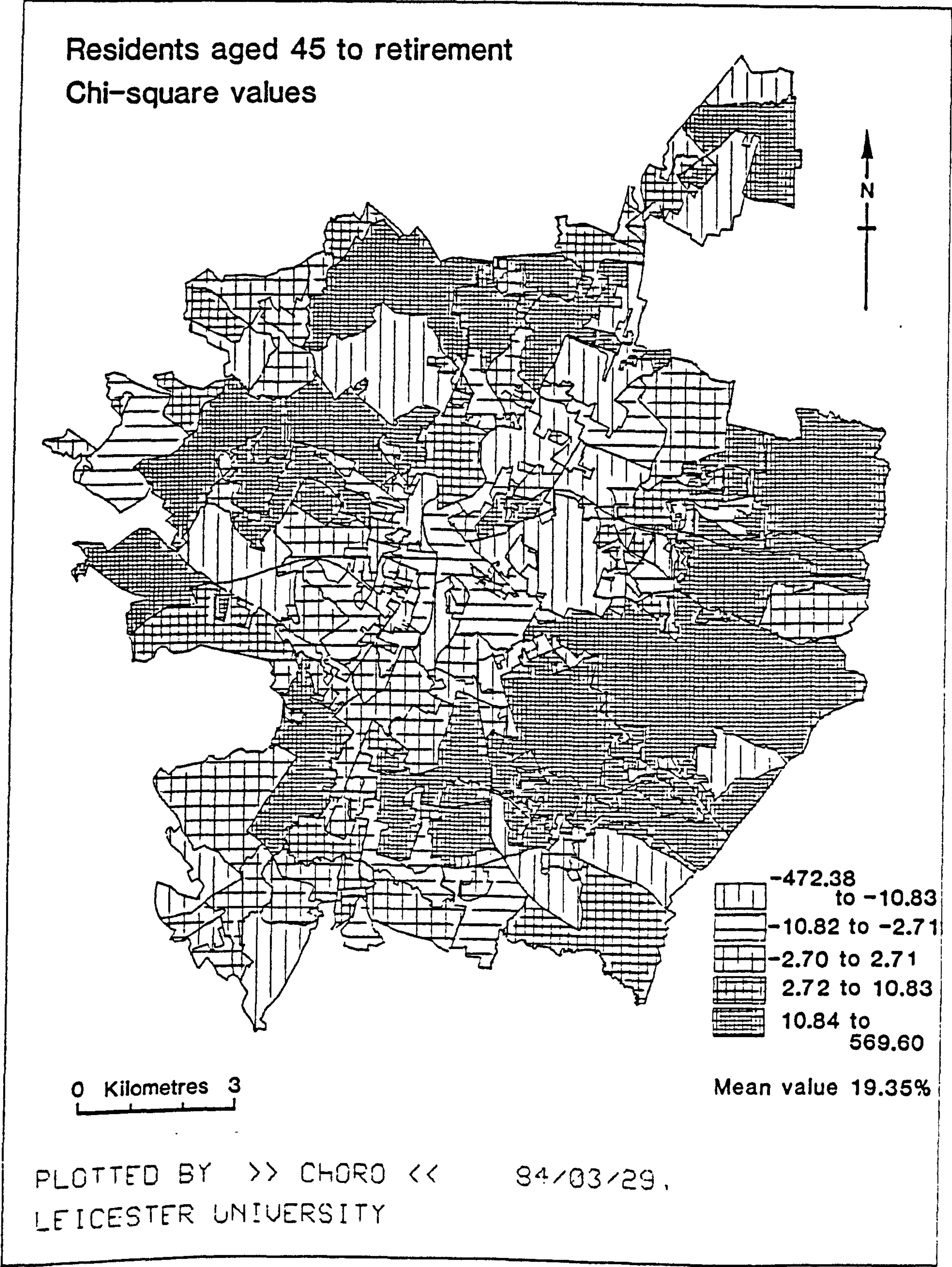
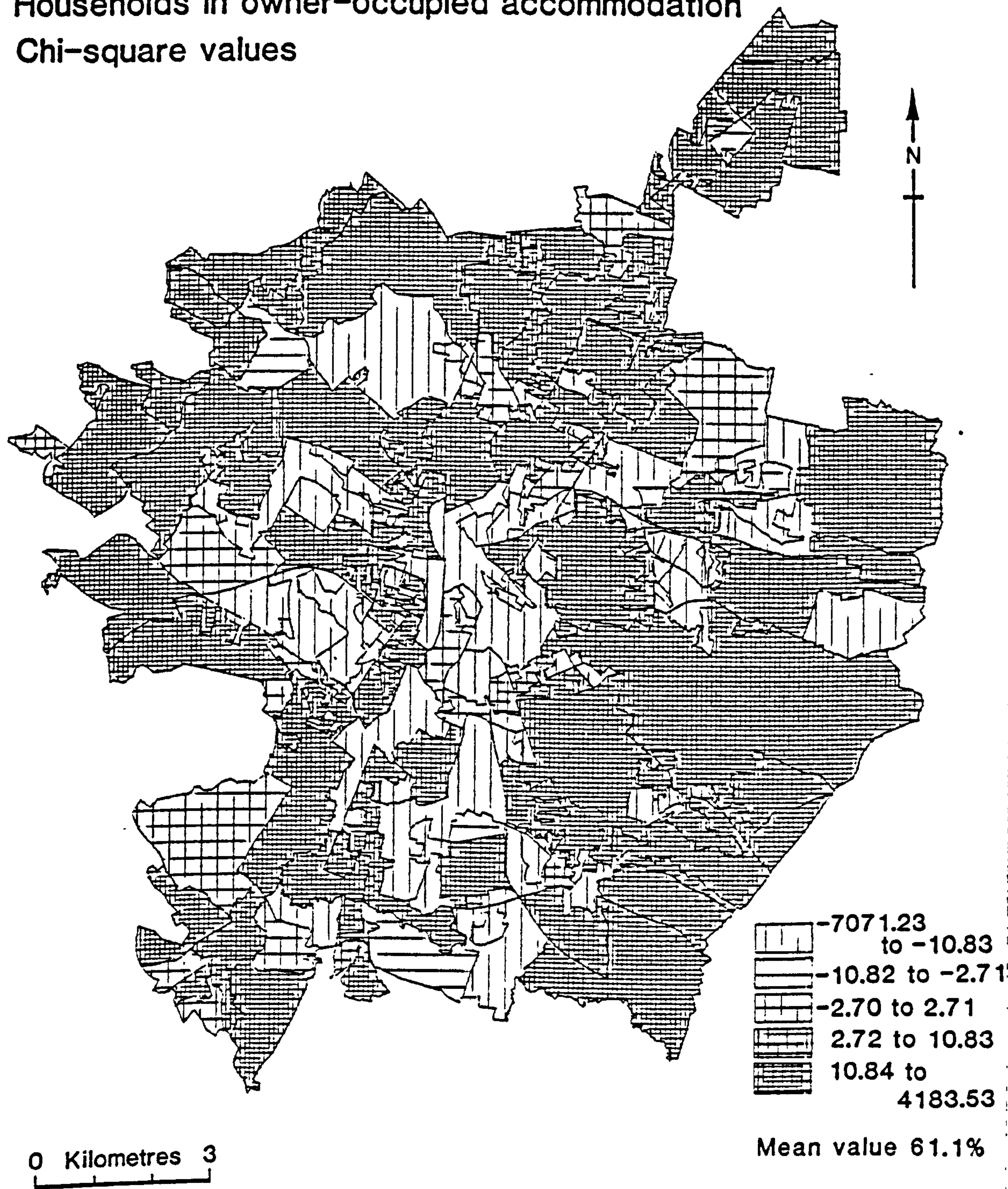


Figure A1.30

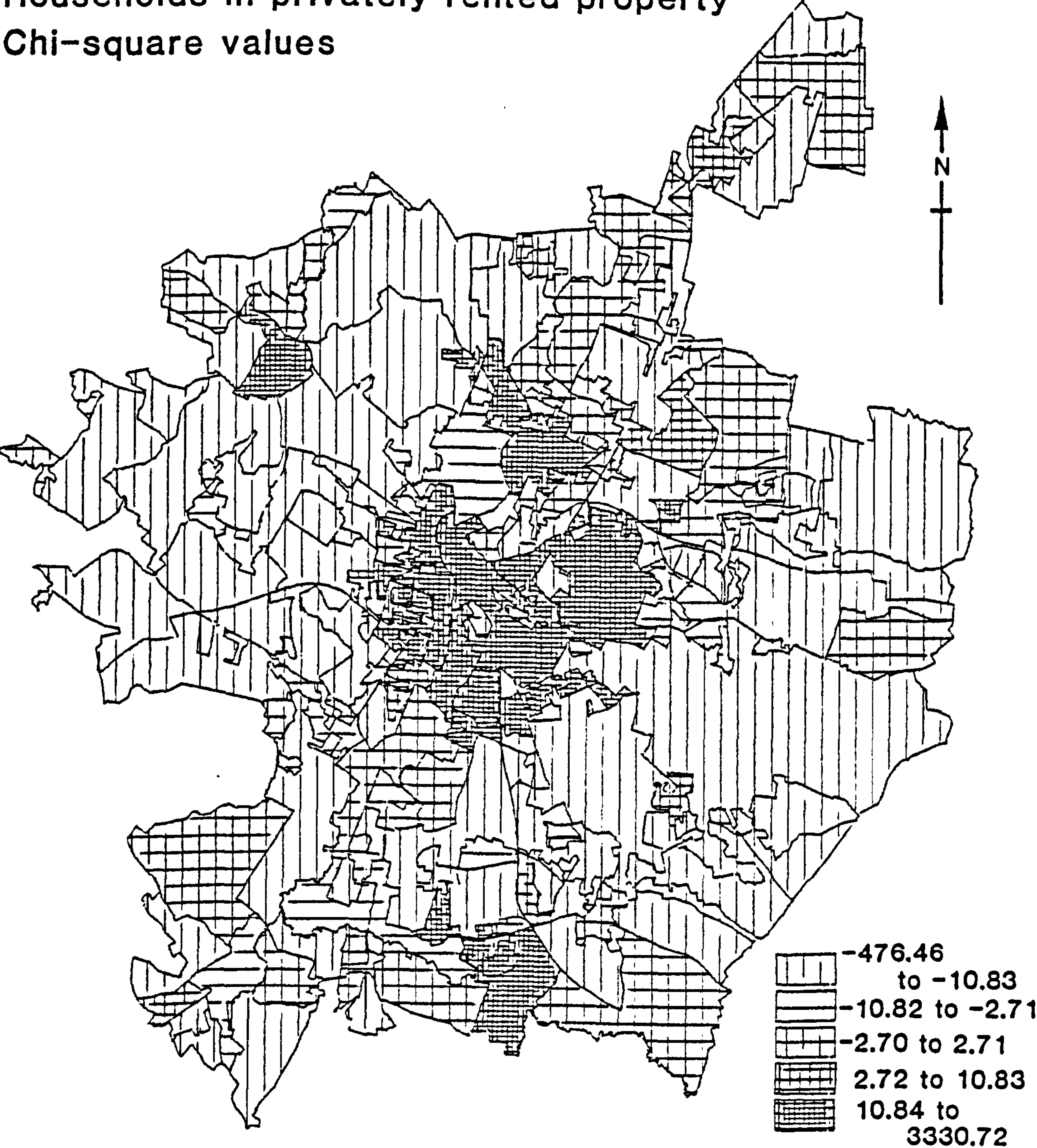
Households in owner-occupied accommodation
Chi-square values



PLOTTED BY >> CHORO << 84/03/29.
LEICESTER UNIVERSITY

Figure A1.31

Households in privately rented property
Chi-square values



0 Kilometres 3

PLOTTED BY >> CHORO << S4/03/29.
LEICESTER UNIVERSITY

Table A1.6 Advantages and disadvantages of the Chi-square statistic for mapping

Advantages:

- Directly comparable across all types of variables.
- values are related to a global mean and incorporate both absolute and relative deviation.
- incorporates a 'weighting' for the size of each area, thereby facilitating comparison of unevenly sized areas
- emphasises significantly low and high values (although its inability to distinguish exceptionally high and low values must be seen as a disadvantage).
- permits a probabilistic (arguably more realistic) approach to mapping.
- it is not based purely on a ratio value and therefore it avoids problems inherent to these (e.g. Visvalingam and Dewdney (1977) have found the chi-square statistic to be more robust than ratio values when the size of the spatial unit of aggregation is changed).

Disadvantages:

- does not have the immediate acceptance and understanding of the ratio or absolute value, the legend may require explanation.
- can only be used for dichotomous classifications - those with and without a characteristic.
- unreliable when the expected number of occurrences in an area is less than 5.

readability and aesthetic value.

The most striking feature of Figure A1.30 showing 'Households in owner-occupied accommodation' is its almost total dominance by the two extreme classes. This reflects quite strongly the original 'U' shaped distribution of the ratio value. It was previously suggested that the arithmetic progression produced the most useful map of ratio values (see Figure A1.8), the chi-square map appears to be very similar, although it shows greater contrast between the areas.

Figure A1.31 very clearly identifies areas with a high proportion of 'Households in privately rented property', it is less successful at distinguishing areas with extremely low proportions since so many are members of the lowest category. It may however be argued that this is unimportant since it correctly emphasises how many areas have fewer privately rented households than expected. Compared with the earlier maps (see Figures A1.14 to A1.19) the map of chi-square values has much greater visual clarity. It depends on the user's viewpoint how much importance is attached to the dominance by the lowest class.

Table A1.6 lists the advantages and disadvantages of using the signed chi-square measure for mapping. These may be compared with those for the ratio values in Table A1.2.

A1.7 Conclusion: evaluation and recommendations.

The first part of this appendix critically evaluated the use of both the choropleth and proportional symbol styles of map for displaying census data. It was shown that the statistical distribution of the data values ultimately determined the effectiveness of the map, subject to certain operational decisions.

Two major projects were undertaken; the first assessed the

usefulness of different systems for choosing class intervals for choropleth style maps, the second examined the effects of data transformation using proportional symbol maps. While the former was able to produce a set of maps among which at least one was satisfactory according to the stated criteria for judgement, the latter appeared to be very unsuitable and therefore was considered no further.

The main problem with the choropleth maps was that no single class interval system proved to be effective for the three statistical distributions considered. It was therefore decided to investigate a further possibility - the chi-square statistic. This had been used to some effect in previous work on census data and it was therefore not surprising that the resultant maps in this project were highly satisfactory. Despite a few disadvantages, the chi-square maps fulfilled almost all of the original criteria. That is, they are directly comparable across different variables, class intervals may be chosen to be meaningful to the data values, both the readability and the general appearance of the maps are good, and both significantly high and low data values can be clearly identified.

It was decided that the chi-square value would be used for mapping the remainder of the census variables. The three disadvantages listed in Table A1.6 will be overcome as follows.

First, the key will be labelled 'expected', 'less than expected' and 'more than expected' rather than with the true class limits. This is imprecise but it conveys the correct general impression. Ratio values will still be available for describing significantly high and low areas.

The second disadvantage does not apply since each of the census variables can be classified in the appropriate manner.

Finally, although chi-square is unreliable for small counts, it is to be hoped that following the clustering and aggregation processes few areas will have unsuitably small populations. In the conventional use of the chi-square statistic some smaller classes are generally permitted if care is taken at the interpretation stage.

Appendix 2. Computer Programming

A2.1 Introduction

In addition to the many software packages implemented in the course of this project, much bespoke programming was also involved. Most of this work entailed fairly straightforward data manipulation as necessitated by the different packages, however some of the specially designed programs performed more complicated data handling. These programs are documented below.

A2.2 Programs for data preparation and manipulation

A2.2.1 (i) Program MPPOLX

(ii) Extracts polygons from a digitiser file

(iii) Input file

a) contains digitised coordinates of boundary segments

Output file

a) contains listed segment names defining polygons

(iv) Program description

a) Main program.

Calls subroutines DIFFER, ANGCLC and SETUP. Controls main processing. Reads digitiser file and for each segment the following information is put into a table: start and end nodes, angles of vectors immediately approaching these nodes, a flag to indicate whether this segment has been used yet to make up a polygon (all but the boundary segments will form the side of two polygons), and the middle character of the segment name. This table is then manipulated to produce lists of segment names defining polygons, node numbers and vector angles are used for this (see Notes below).

b) Subroutine SETUP

Enables user specification of input and output files.

c) Subroutine DIFFER

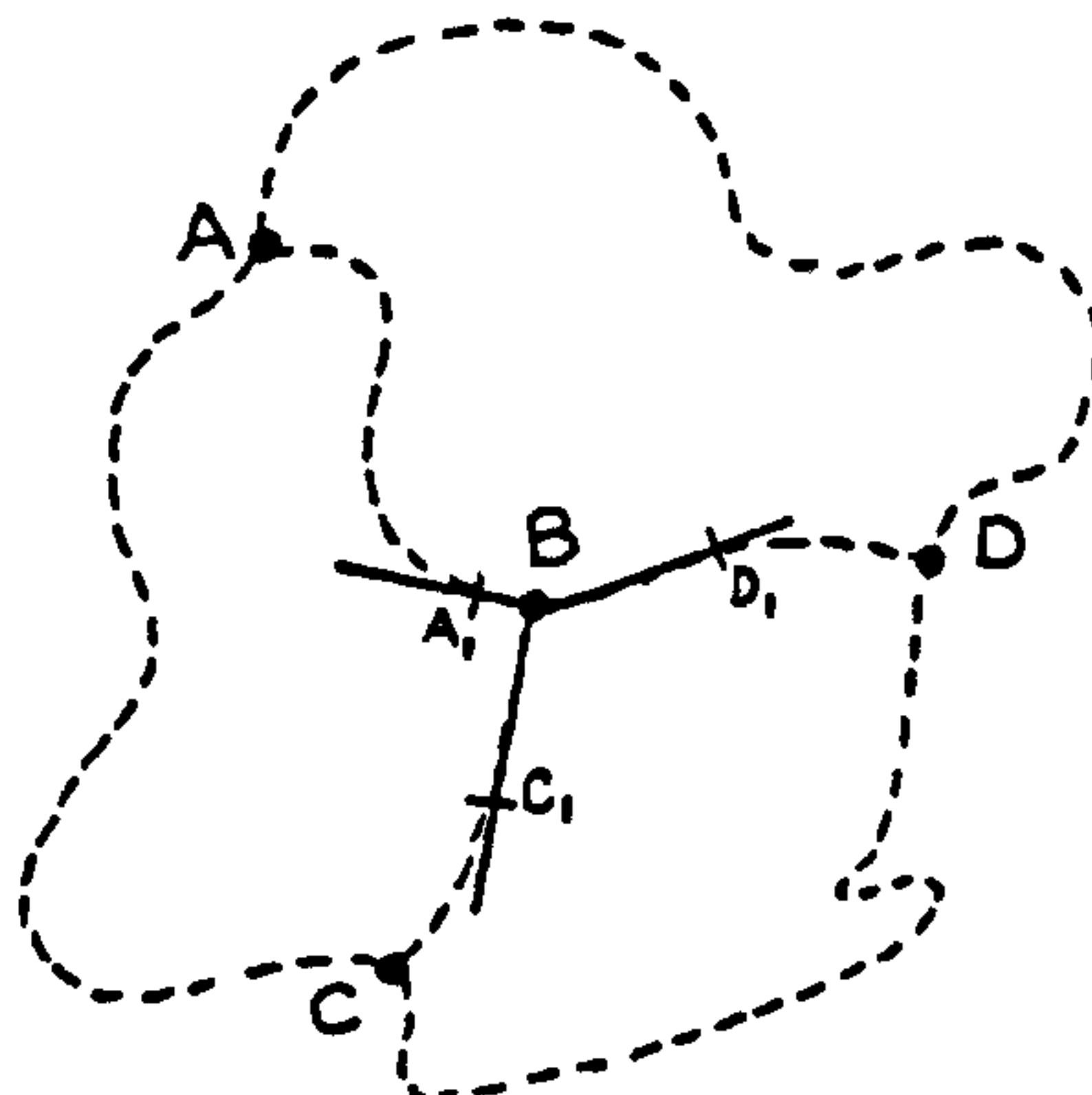
Compares successive values of x and y coordinates to eliminate points located very close together from calculation of the vector angle.

d) Subroutine ANGCLC

Calculates the angle of the vector at each node, giving the angle in radians.

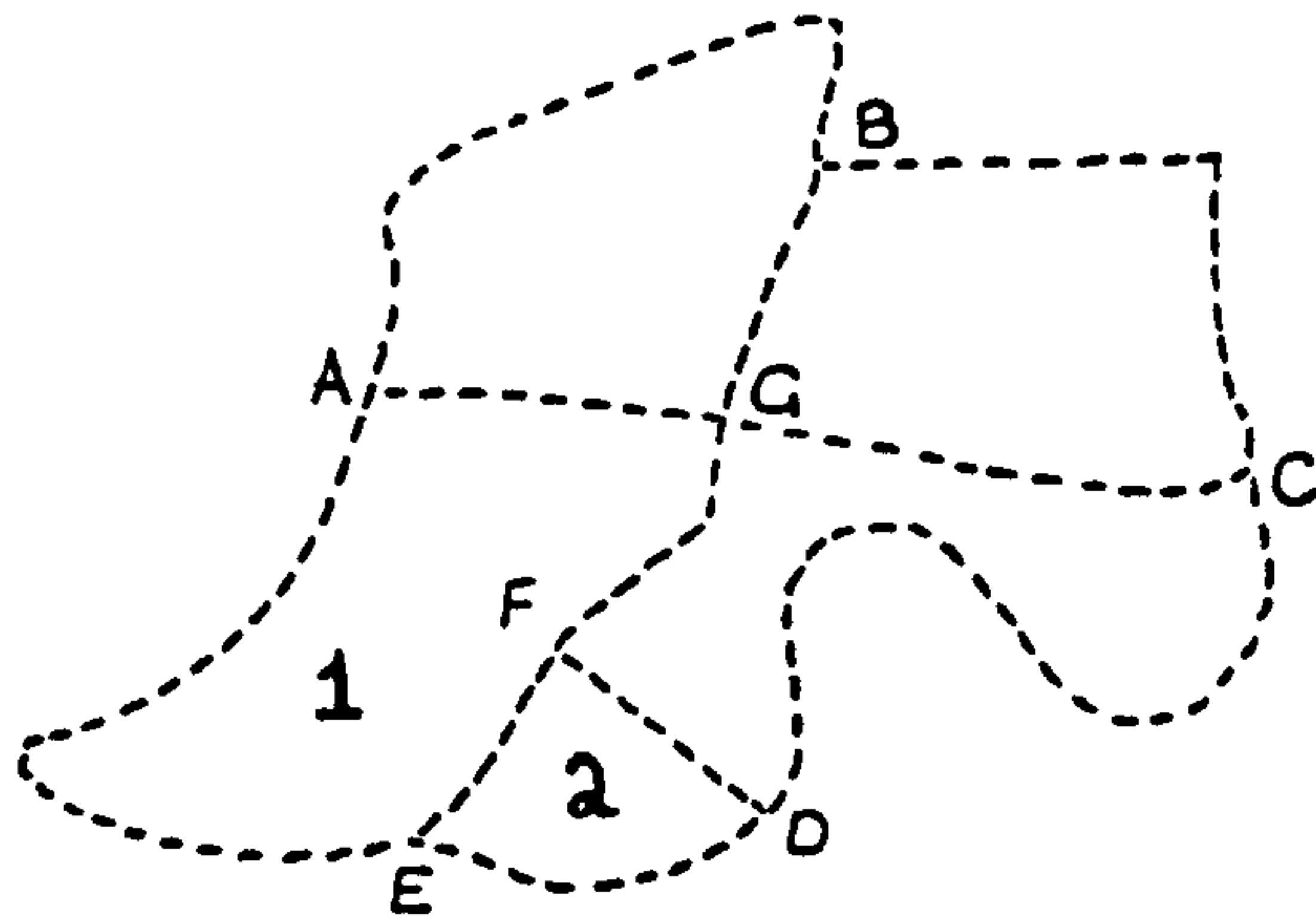
(v) Notes

- a) The basic algorithm: a polygon may be defined as a list of segments given in a clockwise direction if the difference between the vector angles approaching and leaving the node is at a maximum. i.e. for the following node:



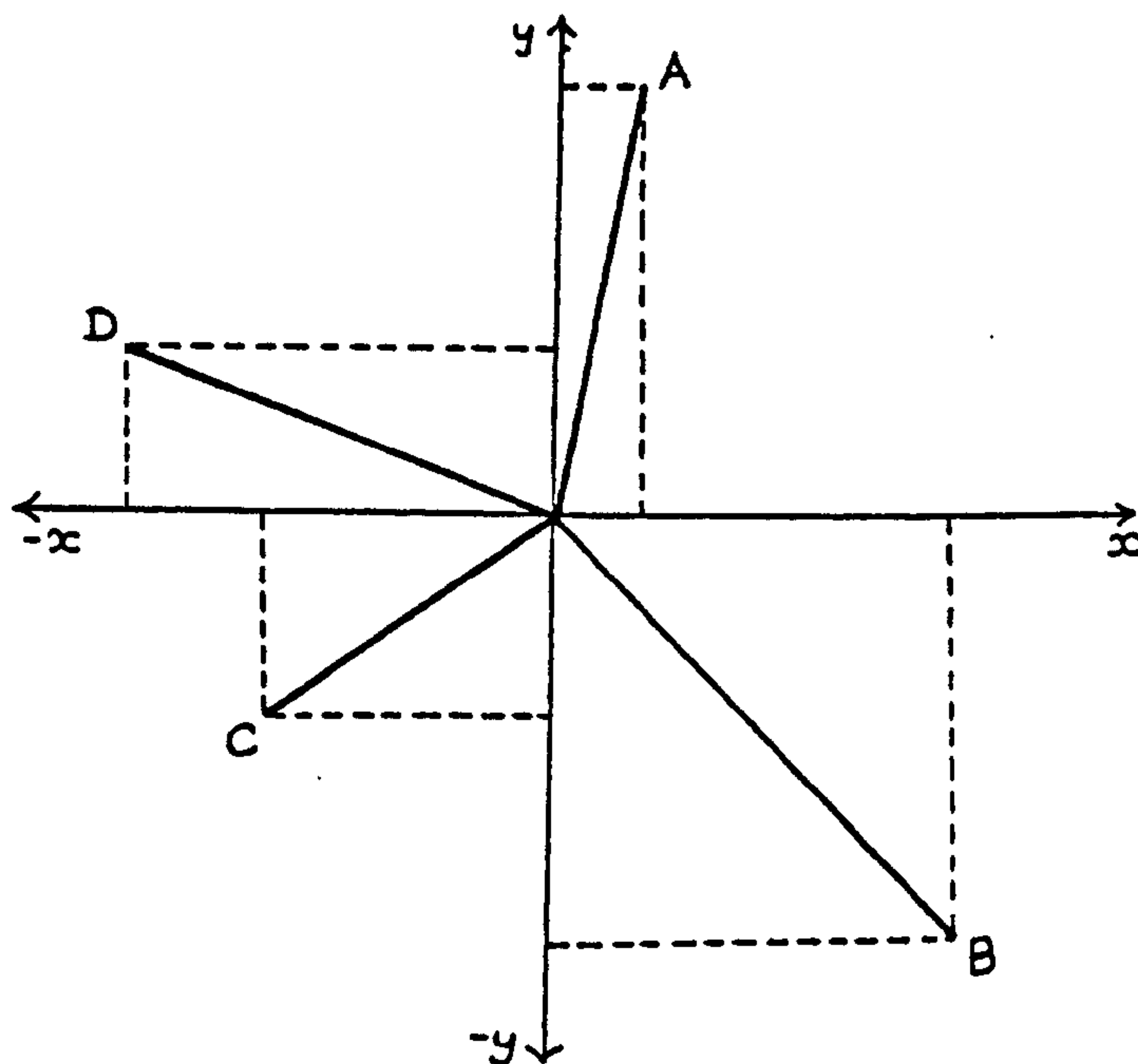
Given a vector to B from A (i.e. A_1B), the program will correctly identify $B+C$ as the next segment in polygon 1 since the angle between A_1B and C_1B is greater (in a clockwise direction) than the angle between A_1B and D_1B .

- b) The exception to this rule occurs when a segment on the outer boundary of the data is found. Such a segment will form the edge of one polygon as well as part of the outer boundary of all polygons in the data set. This is shown below.



Segment E+F forms an edge to polygons 1 and 2 but segment E+D is only part of polygon 2. Since the algorithm requires that each segment be used twice, the program causes the outer boundary to be specified in an anticlockwise direction.

- c) Vector angles are calculated as follows. Assume that in all cases the vector starts at the intersection of the segments (at point x_1y_1)



In all cases: start = x_1y_1 end = $x_2 y_2$

For vectors in the NE quadrant: (e.g. A)

$$\text{Angle} = \text{ATAN} \left(\frac{x_2 - x_1}{y_2 - y_1} \right)$$

PROGRAM PCLYEX(TAPE1,TAPE2,TAPE3,OUTPUT)

```

C
C PROGRAM EXTRACTS POLYGONS FROM A DIGITISER FILE
C
C THE ALGORITHM USES THE ANGLE OF THE VECTOR IMMEDIATELY
C APPROACHING EACH NODE TO DRAW UP A CLOCKWISE LIST
C OF SEGMENTS WHICH WILL DEFINE EACH POLYGON.
C AN ANTICLOCKWISE LIST OF THE OUTER BOUNDING SEGMENTS
C IS ALSO PRODUCED.
C
C REQUIREMENT - SECTION NAMES MUST CONSIST OF THE START
C AND END NODES OF THE SECTION DIGITISED I.E.9999+9999
C
C LIST OF VARIABLES:
C ITAB,JTAB=COUNTERS USED FOR DESIGNATING LOCNS IN TABLE
C IEOPOL=END OF POLYGON MARKER IN POLYGON FILE-PCLFIL
C IECS=END OF SECTION MARKER IN DIGITISER FILE-DIGFIL
C INCDEA,INODEB=NODE NUMBERS READ IN FROM SECTION NAME
C           IN DIGITISER FILE
C XSTART,YSTART=COORDINATES OF START OF VECTOR
C XEND,YEND=COORDINATES OF END OF VECTOR
C XDIFF,YDIFF=DISTANCE (MEASURED IN X AND Y UNITS)
C           BETWEEN START AND END OF VECTOR
C ANGLE=VECTOR ANGLE PUT INTO TABLE, MEASURED IN
C           RADIANS IN A CLOCKWISE DIRECTION (NORTH=0.0)
C JEND=LENGTH OF TABLE
C IFLG=INDICATES WHETHER A SEGMENT HAS ALREADY BEEN
C           ATTACHED TO THIS NODE
C ISTNOD,IENNOD=START AND END NODES OF THE SEGMENT BEING
C           SEARCHED FOR
C IUSFLG=INDICATES WHETHER THIS SEGMENT HAS BEEN USED
C           BEFORE, AND HOW
C INCDEO=NODE OF ORIGINAL SEGMENT WHICH WILL BE USED
C           IN TABLE SEARCH
C IPCLST=STARTING NODE OF THE POLYGON AS WRITTEN TO FILE
C ANGLA=ANGLE OF VECTOR AT INCDEO
C INODTS,INODNE=NODES READ SUCCESSIVELY FROM TABLE TO
C           FIND MATCH WITH INCDEO
C ANGLB=ANGLE OF VECTOR AT NODE WHICH IS FOUND TO MATCH
C ANGDF1=DIFFERENCE BETWEEN VALUES OF ANGLA AND THE
C           VECTOR ANGLE ALREADY ATTACHED
C ANGDF2=DIFFERENCE BETWEEN ANGLA AND THE VECTOR NOW
C           BEING EXAMINED
C JPCSN=HOLDS THE ROW NUMBER IN WHICH THE ATTACHED
C           SEGMENT CAN BE FOUND
C JUSFLG=INDICATES WHETHER SEGMENT HAS BEEN ACCESSED
C           BEFORE - USED TO UPDATE THE VALUE IN THE TABLE
C INODA,INODB=NODE NUMBERS WRITTEN TO POLYGON FILE,
C           USED TO COMPARE WITH IPCLST TO DETERMINE
C           WHETHER THE POLYGON IS COMPLETE
C IRTN=RETURN CODE FROM SUBROUTINE DIFFER, RETURNED AS 1
C           IF THE DIFFERENCE IS NOT SIGNIFICANT
C JOINER=CHARACTER JOINING THE TWO NODES IN THE SEGMENT
C           NAME, USUALLY HAS THE VALUE '+'. USED FOR
C           SETTING UP TABLE
C IJCIN,JOIN=AS JOINER, BUT USED FOR PROCESSING IN
C           EXTRACTING FROM TABLE AND WRITING TO FILE
C ICHANL=HOLDS THE CHANNEL NUMBER FOR THE OUTPUT CHANNEL
C           WHEN CALLING SUBROUTINE WRITEF
C
C CHARACTER*1 IECS,IECPCL,JOINER,JCIN
C DIMENSION ITABL1 (4,1700), TABLE2 (2,1700)

```



```

C
C INITIALISATION OF VARIABLES
C
      ITAB=0
      JTAB=0
      IEOPCL='*'
      PI=3.14159
      ICHANL=2
      IPLCNT=0
      KOUNT=0

C
C FIRST STAGE OF PROCESSING IS TO SET UP THE TABLES.
C ITABL1:
C EACH ROW WILL HAVE 4 COLUMNS CONTAINING THE FOLLOWING
C INFORMATION ABOUT THE SECTION:
C COL 1: STARTING NODE
C COL 2: END NODE
C COL 3: FLAG TO SHOW IF THIS SEGMENT HAS ALREADY BEEN
C        USED IN A POLYGON
C        FLAG = 1 IF THE SECTION HAS BEEN USED IN THE
C                DIRECTION IT IS HELD ON THE DIGITISER
C                FILE (I.E. START AT START NODE)
C        FLAG = 2 IF THE SECTION HAS BEEN USED ONLY IN
C                THE OPPOSITE DIRECTION
C        FLAG = 3 IF THE SECTION HAS BEEN USED BOTH WAYS
C                - IT CAN THEREFORE NOT BE USED AGAIN
C COL 4: HOLDS A REPRESENTATION OF THE CHARACTER WHICH
C        JOINS THE TWO NODES IN A SECTION NAME
C TABLE2:
C COL 1: ANGLE IN RADIANs AT WHICH THE SEGMENT LEAVES
C        THE STARTING NODE
C COL 2: ANGLE IN RADIANs AT WHICH THE SEGMENT APPROACHES
C        THE END NODE
C FIRSTLY, FIND THE END OF THE INITIAL SECTION
C
100  READ (1,1000,END=15,ERR=100)IEGS
      IF (IEGS.NE.'-') GOTO 100

C
C READ THE NEXT SECTION NAME
C
      READ (1,1000)
      READ (1,2000,END=10) INODEA,JOINER,INODEB

C
C PUT THE SECTION NAME INTO THE TABLE
C
      ITAB=1
      JTAB=JTAB+1
      ITABL1 (ITAB,JTAB) = INODEA
      ITAB=2
      ITABL1 (ITAB,JTAB) = INODEB

C
C READ THE FIRST TWO PAIRS OF COORDINATES IN THE SECTION
C AND CHECK THAT THEY ARE SIGNIFICANTLY DIFFERENT
C (THIS IS TO ENSURE THAT RESULTS ARE NOT DISTORTED
C  BECAUSE OF THE WOBBLE FACTOR)
C THEY MUST DIFFER BY AT LEAST A FACTOR OF 0.001 IN BOTH THE
C X AND Y DIRECTION
C IF THEY DO NOT DIFFER BY THIS THE NEXT PAIR OF COORDINATES
C IS READ AND COMPARED. THIS IS REPEATED UNTIL A SIGNIFICANT
C DIFFERENCE IS FOUND
C
      READ (1,4000)X,Y

```

```

      XSTART = X
      YSTART = Y
500  IF (INCDEA.EQ.858.AND.INCDEB.EQ.1115) THEN
      KOUNT=KOUNT+1
      WRITE (3,*) KCLNT,XSTART,YSTART,XEND,YEND
      ENDIF
      READ (1,4000)X,Y
      IF (X.EQ.0) GOTO 550
      XEND = X
      YEND = Y
      IRTN=0
      CALL DIFFER(XSTART,YSTART,XEND,YEND,IRTN)
      IF (IRTN.NE.0)GOTO 500
C
C X=0.0 (I.E.BLANKS READ IN E FORMAT) WHEN END OF SECTION IS
C REACHED, THEREFORE USE LAST VALUES OF XEND AND YEND
C TO CALCULATE THE VECTOR ANGLE
C
550  CONTINUE
C
C CALCULATE THE ANGLE OF THIS VECTOR AND PUT IT IN THE TABLE
C
400  CALL ANGCLC (XSTART,YSTART,XEND,YEND,ANGLE)
      ITAB = 1
      TABLE2 (ITAB,JTAB) = ANGLE
C
C GO TO THE BOTTOM OF THIS SECTION AND START READING THE FILE
C BACKWARDS TO GET THE ANGLE AT THE OTHER END
C THE 'BACKSPACE' COMMAND SPECIFIES WHICH CHANNEL IS TO BE
C READ BACKWARDS
C
      BACKSPACE 1
600  READ (1,1000,ERR=600)IEOS
      IF (IEOS.NE.'-') GOTC 600
      BACKSPACE 1
      BACKSPACE 1
      READ (1,4000)X,Y
      XSTART=X
      YSTART=Y
      KOUNT=0
121  IF (INDEA.EQ.858.AND.INDEB.EQ.1115) THEN
      KOUNT=KOUNT+1
      WRITE(3,*) KOUNT,XSTART,YSTART,XEND,YEND
      ENDIF
      BACKSPACE 1
      BACKSPACE 1
      READ (1,4000)X,Y
      IF (Y.EQ.0) GOTC 122
      XEND=X
      YEND=Y
      IRTN=0
      CALL DIFFER(XSTART,YSTART,XEND,YEND,IRTN)
      IF (IRTN.NE.0)GOTC 121
122  CONTINUE
C
C CALCULATE THE ANGLE OF THIS END VECTOR AND PUT IT IN THE TABLE
C
      CALL ANGCLC (XSTART,YSTART,XEND,YEND,ANGLE)
      ITAB = 2
      TABLE2 (ITAB,JTAB) = ANGLE
C
C SET THE ACCESS FLAG TO ZERO

```

```

C
      ITAB = 3
      ITABL1 (ITAB,JTAB) = 0
C
C PUT THE CHARACTER JOINING THE TWO NCDES IN THE SEGMENT NAME
C INTO THE TABLE I.E '+'=0.0, 'A'=1.0, 'B'=2.0
      IF (JOINER.EQ.'+') ITABL1(4,JTAB)=0
      IF (JOINER.EQ.'A') ITABL1(4,JTAB)=1
      IF (JOINER.EQ.'B') ITABL1(4,JTAB)=2
      GOTO 100
C
C
C
C SECOND STAGE OF PROCESSING SORTS THE ENTRIES IN THE TABLE
C INTO POLYGONS
C
C JEND HOLDS THE TOTAL NUMBER OF LINES IN THE TABLE
C
      10  JEND=JTAB
      112 JTAB = 0
C
C READ THE FIRST (NEXT) ENTRY TO THE TABLE, CHECKING THAT
C THE SEGMENT HAS NOT BEEN USED TWICE BEFORE. THIS IS
C TAKEN AS THE STARTING SEGMENT OF THE FIRST (NEXT) POLYGON,
C AND ITS 'IUSFLG' IS UPDATED TO ACCOUNT FOR ITS USE
C AS THE FIRST SEGMENT IN A POLYGON
C IF THE END OF THE TABLE IS REACHED BEFORE FINDING
C IUSFLG.NE.3.0 THEN ALL SEGMENTS HAVE BEEN USED TWICE AND
C THE PROGRAM TERMINATES
C
      107 IF (JTAB.EQ.JEND) GOTO 9999
          JTAB = JTAB + 1
          IFLG=0
          IUSFLG = ITABL1 (3,JTAB)
          IF (IUSFLG.EQ.3) GOTO 107
          ITAB = 1
          ISTNOD = ITABL1 (ITAB,JTAB)
          ITAB = 2
          IENNOD = ITABL1 (ITAB,JTAB)
          ITAB = 4
          IJOIN=ITABL1(ITAB,JTAB)
          IF (IUSFLG.EQ.2.0R.IUSFLG.EQ.0) GOTO 108
          IF (IUSFLG.EQ.1.0) GOTO 109
C
      108 INCDEO = IENNOD
          IPCLST = ISTNOD
          ANGLA = TABLE2(2,JTAB)
          IF (IUSFLG.EQ.2.0) ITABL1(3,JTAB)=3
          IF (IUSFLG.EQ.0.0) ITABL1(3,JTAB)=1
          GOTO 110
C
      109 INCDEO = ISTNOD
          IPCLST = IENNOD
          ANGLA = TABLE2 (1,JTAB)
          ITABL1(3,JTAB)=3
C
C WRITE THIS SEGMENT INTO THE POLYGON FILE (INCLUDING
C THE CORRECT CHARACTER BETWEEN THE NODE NUMBERS)
C
      110 IF (IJOIN.EQ.0.0) JOIN='+'
          IF (IJOIN.EQ.1.0) JOIN='A'

```



```

      IF (IJOIN.EQ.2.0) JOIN='B'
      CALL WRITEFC(ISTNOD,JOIN,IENNOD,ICHANL)
C
C READ DOWN THE TABLE SEARCHING FOR A NODE WHICH
C MATCHES INODEO
C
      300 IF (JTAB.EQ.JEND) GOTO 103
          JTAB = JTAB + 1
C
C CHECK THAT THE SEGMENT FOUND HAS NOT BEEN USED TWICE ALREADY
C
      IUSFLG = ITABL1(3,JTAB)
      IF (IUSFLG.EQ.3) GOTO 300
C
C EXAMINE THE MATCHING NODE
C
      ITAB = 1
      INODTS = ITABL1(ITAB,JTAB)
      ITAB = 2
      INODNE = ITABL1(ITAB,JTAB)
C
C CHECK THAT FIRST NODE NOT PREVIOUSLY USED
C
      IF (IUSFLG.EQ.1) GOTO 104
      IF (INODEO.EQ.INODTS) GOTO 700
C
C CHECK THAT 2ND NODE NOT PREVIOUSLY USED
C
      104 IF (IUSFLG.EQ.2) GOTO 300
          IF (INODEO.EQ.INODNE) GOTO 900
          GOTO 300
C
C READ THE ANGLE OF THE APPROPRIATE NODE FROM THE TABLE
C
      700 ITAB = 1
          ANGLB = TABLE2(ITAB,JTAB)
          IFLG = IFLG + 1
          GOTO 101
      900 ITAB = 2
          ANGLB = TABLE2(ITAB,JTAB)
          IFLG = IFLG + 1
C
C CHECK IF THERE HAS BEEN A PREVIOUS SEGMENT JOINED TO
C THIS NODE, IF NOT THIS ONE MAY BE JOINED NOW
C
      101 WRITE(3,1111) INODTS,INODNE,ANGLA,ANGLB
1111  FORMAT (1X,I4,1X,I4,1X,F8.6,1X,F8.6)
      IF (IFLG.GT.1) GOTO 102
C
C CALCULATE THE DIFFERENCE BETWEEN THE ORIGINAL SEGMENT
C AND THAT NOW BEING ATTACHED
C
      IF (ANGLB.GT.ANGLA) ANGDF1 = ANGLB - ANGLA
      IF (ANGLA.GT.ANGLB) ANGDF1 = (2 * PI) - (ANGLA-ANGLB)
      IF (ANGLA.EQ.ANGLB) ANGDF1 = 0.0
      JPOSN = JTAB
      GOTO 300
C
C IF THERE IS A PREVIOUS SEGMENT ITS ANGLE TO THE ORIGINAL
C (ANGDF1) WILL HAVE TO BE COMPARED WITH THE ANGLE OF
C THIS NEW SEGMENT (ANGDF2)
C

```

```

102  IF (ANGL5.GT.ANGLA) ANGDF2 = ANGL3-ANGLA
      IF (ANGLA.GT.ANGL3) ANGDF2 = (2 * PI) - (ANGLA-ANGL3)
      IF (ANGLA.EQ.ANGL3) ANGDF2 = 0.0
      IF (ANGDF2.LE.ANGDF1) GOTO 900
C
C IF THE NEW ANGLE IS LARGER THEN THE NEW SEGMENT REPLACES
C THE PREVIOUS SEGMENT, EITHER WAY THE SEARCH CONTINUES
C
      ANGDF1 = ANGDF2
      JPSN = JTAB
      GOTO 800
C
C PROCESSING WHEN THE END OF THE TABLE IS REACHED:
C
C FIRSTLY, WRITE THE CORRECT SEGMENT AWAY TO FILE
C (INCLUDING THE CORRECT CHARACTER BETWEEN THE NODE
C NUMBERS
C
103  IJOIN=ITABL1(4,JPCSN)
      IF (IJOIN.EQ.0) JCIN='+'
      IF (IJOIN.EQ.1) JCIN='A'
      IF (IJOIN.EQ.2) JCIN='B'
      ISTNOD = ITABL1(1,JPSN)
      IENNOD = ITABL1(2,JPSN)
      CALL WRITEF (ISTNOD,JCIN,IENNOD,ICHANL)
C
C NEXT, ENTER APPROPRIATE FLAG IN TABLE TO SHOW THIS
C SEGMENT HAS BEEN USED IN THIS DIRECTION
C
      IF (ITABL1(1,JPSN).EQ.INODE0) ITAB = 1
      IF (ITABL1(2,JPSN).EQ.INODE0) ITAB = 2
      JUSFLG = ITABL1(3,JPSN)
      IF (JUSFLG.NE.0) GOTO 105
      IF (ITAB.EQ.1) ITABL1(3,JPSN)=1
      IF (ITAB.EQ.2) ITABL1(3,JPSN)=2
      GOTO 106
105  ITABL1(3,JPSN)=3
C
C STORE THE NODES OF THE SEGMENT WRITTEN TO FILE TO
C TEST WHETHER THE POLYGON IS COMPLETE
C
106  INODA = ITABL1(1,JPSN)
      INODB = ITABL1(2,JPSN)
      IF (INODA.NE.IPOLST.AND.INODB.NE.IPOLST)GOTO 111
      WRITE (2,1000)IEOPCL
C
C WRITE MESSAGE TO THE TERMINAL WHEN 100 POLYGONS HAVE
C BEEN PROCESSED
C
      IPLCNT=IPLCNT+1
      IF(IPLCNT.EQ.100)GOTO 333
      GOTO 112
333  WRITE(3,3333)
3333 FORMAT (1H0,' 100 POLYGONS PROCESSED ')
      IPLCNT=0
      GOTO 112
C
C IF A NEW POLYGON IS NOT BEING STARTED THEN SET UP INODE0
C AND ANGLA FROM THE MOST RECENT SEGMENT ATTACHED. (ITAB
C AT THIS STAGE HOLDS THE POSITION OF THE NODE JUST
C ATTACHED, SO WE WANT TO PROCESS THE OTHER ONE).
111  IF (ITAB.EQ.1) IPCSN = 2

```

```

      IF (ITAB.EQ.2) IPCSN = 1
      INDEO = ITABL1(IPCSN,JPOSN)
      ANGLA = TABLE2(IPCSN,JPCSN)
      JTAB = 0
      IFLG = 0
      GOTO 800
C
C ERROR MESSAGES
C
      15  WRITE (5,6000)
          GOTO 9999
C
C FORMAT STATEMENTS
C
      1000  FORMAT (A1)
      2000  FORMAT (I4,A1,I4)
      4000  FORMAT (5X,E13.7,2X,E13.7)
      5000  FORMAT (I4,A1,I4)
      6000  FORMAT (1H0,'END OF FILE BEFORE END OF INITIAL SECTION')
C
C FINISH
C
      9999  STOP
          END
C
C
C
      SUBROUTINE DIFFER(XSTART,YSTART,XEND,YEND,IRTN)
C
C SUBROUTINE CHECKS THAT THERE IS AT LEAST A DIFFERENCE OF
C (0.01,0.01) BETWEEN (XSTART,YSTART) AND (XEND,YEND)
C
      XDIF = XSTART - XEND
      YDIF = YSTART - YEND
      IF (XDIF.GE.0.0.AND.XDIF.LE.0.01.OR.XDIF.LE.0.0
+       .AND.XDIF.GE.-0.01) IRTN=1
      IF (YDIF.GE.0.0.AND.YDIF.LE.0.01.OR.YDIF.LE.0.0
+       .AND.YDIF.GE.-0.01) IRTN=1
      RETURN
      END
C
C
C
      SUBROUTINE ANGCLC(XSTART,YSTART,XEND,YEND,ANGLE)
C
C SUBROUTINE CALCULATES THE ANGLE AT WHICH THE
C VECTOR APPROACHES THE END POINT
C
      PI=3.14159
C
C FIRSTLY ELIMINATE VECTORS LYING ON THE CARDINAL POINTS
C
      IF (YEND.EQ.YSTART.AND.XEND.GT.XSTART) ANGLE=0.5*PI
      IF (XEND.EQ.XSTART.AND.YEND.LT.YSTART) ANGLE=PI
      IF (YEND.EQ.YSTART.AND.XEND.LT.XSTART) ANGLE=1.5*PI
      IF (XEND.EQ.XSTART.AND.YEND.GT.YSTART) ANGLE=0.0
C
C CALCULATE ANGLE FOR VECTORS IN NE QUADRANT
C
      IF (XEND.GT.XSTART.AND.YEND.GT.YSTART)
+       ANGLE = ATAN((XEND-XSTART)/(YEND-YSTART))

```



```

C
C CALCULATE ANGLE FOR VECTORS IN SE QUADRANT
C
      IF (XEND.GT.XSTART.AND.YEND.LT.YSTART)
      +   ANGLE = (0.5*PI) + ATAN((YSTART-YEND)/(XEND-XSTART))
C
C CALCULATE ANGLE FOR VECTORS IN SW QUADRANT
C
      IF (XEND.LT.XSTART.AND.YEND.LT.YSTART)
      +   ANGLE = PI + ATAN((XSTART-XEND)/(YSTART-YEND))
C
C CALCULATE ANGLE FOR VECTORS IN NW QUADRANT
C
      IF (XEND.LT.XSTART.AND.YEND.GT.YSTART)
      +   ANGLE = (1.5*PI) + ATAN((YEND-YSTART)/(XSTART-XEND))
C
      RETURN
      END
C
C
C
C
C
      SUBROUTINE WRITEF (ISTNOD,JOIN,IENNOD,ICHANL)
C
C ENSURES THAT LEADING SPACES IN THE SEGMENT NAMES ARE NOT
C WRITTEN TO THE OUTPUT FILE AS SPACES
C
      CHARACTER*1 JCIN,ICNE,IZERO
      IZERO='0'
C
C FIRSTLY ESTABLISH HOW MANY, AND WHERE, ARE THE LEADING ZEROS
C
      IF (ISTNOD.LT.10.AND.IENNOD.LT.10)GOTO 1001
      IF (ISTNOD.LT.10.AND.IENNOD.LT.100)GOTO 1002
      IF (ISTNOD.LT.10.AND.IENNOD.LT.1000)GOTO 1003
      IF (ISTNOD.LT.10.AND.IENNOD.GE.1000)GOTO 1004
C
      IF (ISTNOD.LT.100.AND.IENNOD.LT.10)GOTO 1005
      IF (ISTNOD.LT.100.AND.IENNOD.LT.100)GOTO 1006
      IF (ISTNOD.LT.100.AND.IENNOD.LT.1000)GOTO 1007
      IF (ISTNOD.LT.100.AND.IENNOD.GE.1000)GOTO 1008
C
      IF (ISTNOD.LT.1000.AND.IENNOD.LT.10)GOTO 1009
      IF (ISTNOD.LT.1000.AND.IENNOD.LT.100)GOTO 1010
      IF (ISTNOD.LT.1000.AND.IENNOD.LT.1000)GOTO 1011
      IF (ISTNOD.LT.1000.AND.IENNOD.GE.1000)GOTO 1012
C
      IF (ISTNOD.GE.1000.AND.IENNOD.LT.10)GOTO 1013
      IF (ISTNOD.GE.1000.AND.IENNOD.LT.100)GOTO 1014
      IF (ISTNOD.GE.1000.AND.IENNOD.LT.1000)GOTO 1015
      IF (ISTNOD.GE.1000.AND.IENNOD.GE.1000)GOTO 1016
C
C ADJUST THE 'WRITE' STATEMENT SO ALL ZEROS ARE WRITTEN TO FILE
C
C FORMAT 000N+000N
C
1001  WRITE(ICHANL,5001)IZERO,IZERO,IZERO,ISTNOD,JOIN,
      +   IZERO,IZERO,IZERO,IENNOD
5001  FORMAT (3A1,I1,A1,3A1,I1)
      GOTO 1999
C

```

```

C FORMAT 000N+00NN
C
1002 WRITE (ICCHANL,5002) IZERO, IZERO, IZERO, ISTNCD, JOIN,
+ IZERO, IZERO, IENNCD
5002 FORMAT(3A1, I1, A1, 2A1, I2)
GOTO 1999
C
C FORMAT 000N+0NNN
C
1003 WRITE(ICCHANL,5003) IZERO, IZERO, IZERO, ISTNOC, JOIN,
+ IZERO, IENNOC
5003 FORMAT (3A1, I1, A1, A1, I3)
GOTO 1999
C
C FORMAT 000N+NNNN
C
1004 WRITE(ICCHANL,5004) IZERO, IZERO, IZERO, ISTNOD, JCIN, IENNOD
5004 FORMAT(3A1, I1, A1, I4)
GOTO 1999
C
C FORMAT 00NN+000N
C
1005 WRITE(ICCHANL,5005) IZERO, IZERO, ISTNCD, JOIN,
+ IZERO, IZERO, IZERO, IENNOC
5005 FORMAT(2A1, I2, A1, 3A1, I1)
GOTO 1999
C
C FORMAT 00NN+00NN
C
1006 WRITE(ICCHANL,5006) IZERO, IZERO, ISTNCD, JCIN,
+ IZERO, IZERO, IENNOC
5006 FORMAT(2A1, I2, A1, 2A1, I2)
GOTO 1999
C
C FORMAT 00NN+0NNN
C
1007 WRITE(ICCHANL,5007) IZERO, IZERO, ISTNCD, JOIN, IZERO, IENNOC
5007 FORMAT(2A1, I2, A1, A1, I3)
GOTO 1999
C
C FORMAT 00NN+NNNN
C
1008 WRITE (ICCHANL,5008) IZERO, IZERO, ISTNOD, JOIN, IENNOC
5008 FORMAT(2A1, I2, A1, I4)
GOTO 1999
C
C FORMAT 0NNN+000N
C
1009 WRITE (ICCHANL,5009) IZERO, ISTNOD, JOIN,
+ IZERO, IZERO, IZERO, IENNOC
5009 FORMAT (A1, I3, A1, 3A1, I1)
GOTO 1999
C
C FORMAT 0NNN+00NN
C
1010 WRITE(ICCHANL,5010) IZERO, ISTNOD, JOIN, IZERO, IZERO, IENNOC
5010 FORMAT(A1, I3, A1, 2A1, I2)
GOTO 1999
C
C FORMAT 0NNN+0NNN
C
1011 WRITE (ICCHANL,5011) IZERO, ISTNOD, JOIN, IZERO, IENNOC

```

```

5011  FORMAT(A1,I3,A1,A1,I3)
      GOTO 1999
C
C  FORMAT 0NNN+NNNN
C
1012  WRITE(ICHANL,5012) IZERO,ISTNOD,JOIN,IENNOD
5012  FORMAT (A1,I3,A1,I4)
      GOTO 1999
C
C  FORMAT NNNN+000N
C
1013  WRITE (ICHANL,5013) ISTNOD,JOIN,IZERO,IZERC,IZERO,IENNOD
5013  FORMAT(I4,A1,3A1,I1)
      GOTO 1999
C
C  FORMAT NNNN+00NN
C
1014  WRITE(ICHANL,5014) ISTNOD,JOIN,IZERC,IZERO,IENNOD
5014  FORMAT (I4,A1,2A1,I2)
      GOTO 1999
C
C  FORMAT NNNN+0NNN
C
1015  WRITE(ICHANL,5015) ISTNOD,JOIN,IZERO,IENNOD
5015  FORMAT (I4,A1,A1,I3)
      GOTO 1999
C
C  FORMAT NNNN+NNNN
C
1016  WRITE(ICHANL,5016) ISTNOD,JOIN,IENNOD
5016  FORMAT (I4,A1,I4)
C
1999  RETURN
      END

```


For vectors in the SE quadrant: (e.g. B)

$$\text{Angle} = \frac{1}{2} \pi + \text{ATAN} \left(\frac{y_1 - y_2}{x_2 - x_1} \right)$$

For vectors in the SW quadrant: (e.g. C)

$$\text{Angle} = \pi + \text{ATAN} \left(\frac{x_1 - y_2}{y_1 - y_2} \right)$$

For vectors in the NW quadrant (e.g. D)

$$\text{Angle} = \frac{3}{2} \pi + \text{ATAN} \left(\frac{y_2 - y_1}{x_1 - x_2} \right)$$

Vectors on the cardinal points take the following values:

$$N = 0$$

$$E = \frac{1}{2} \pi$$

$$S = \pi$$

$$W = \frac{3}{2} \pi$$

d) Data requirement -

Each segment must be named in the following format:

9999a9999

where '9999' is the number of the start or end node and is unique

'a' is usually '+', unless two segments join the same two nodes, in which case one will be named '9999A9999' and the other '9999B9999'.

e) Thanks to Dr. A. Matthews of Leicester University Computer Centre.

A2.2.2 (i) Program MPAGGR

(ii) Creates a file containing a list of bounding segments, in clockwise order, to an area (or areas) defined by the user as a list of enumeration districts, i.e. this program aggregates EDs into larger areas.

(iii) Input files:

a) Data file containing a list of ED names followed by the

names of the segments which bound each ED.

- b) Unformatted file containing the digitised coordinates for each segment, each record containing the following information: segment name, number of pairs of coordinates and an array of x and y coordinates.
- c) List of new area names followed by the names of their constituent EDs.

Output files

- a) Final output - a list of area names followed by the names of their boundary segments listed in clockwise order.
- b) Intermediate output - a list of area names followed by all the segment names from all their constituent EDs. Produced by subroutine EDSEGN.
- c) Intermediate output - a list of names followed by their unsorted boundary segments. Produced by subroutine AREABY.

(iv) Program description.

- a) Main program.

Behaves as a 'driver' section of code, calling the subroutines EDSEGN, AREABY and BDYSRT. Sets up the input and output channels.

- b) Subroutine EDSEGN.

Reads input file containing names of new areas and their constituent EDs and compiles a list of all segments in each new area, writes these to output file MPEDSECT.

- c) Subroutine AREABY.

Reads MPEDSECT and extracts a list of actual boundary segments for each area. This is possible if it is

assumed that all but the boundary segments are part of two EDs. Calls subroutine WRITEF. Creates output file MPABDRY.

d) Subroutine BDYSRT.

Takes the unsorted list of segments from MPABDRY and sorts them into a connecting sequence. Calls subroutine DIRCTN. If the segments are not listed in a clockwise order already, they are reversed before being written to output file MPFAGGR.

e) Subroutine DIRCTN

Uses segment coordinates to establish whether areas are defined in a clockwise or anticlockwise direction, returns a flag to BDYSRT depending on the result. Calls subroutines COORDS and ALDIRN.

f) Subroutine WRITEF

Fills leading spaces in segment names with zeros to maintain '9999+9999' format.

g) Subroutine COORDS.

Accepts an array of segment coordinates and if necessary reverses their order to ensure that coordinates follow consecutively from one segment to the next.

h) Subroutine ALDIRN

Uses a G.A.G. algorithm to check the direction in which an area's boundary is specified (see Notes below).

(v) Notes.

a) The G.A.G. algorithm for determining the direction of a polygon's boundary uses the area of that polygon as calculated from successive values of x and y coordinates. Thus, where A is the area:

$$A = \frac{1}{2} \sum (x_{i-1} - x_i) * (y_{i-1} + y_i)$$

Program MPAGGR

```

PROGRAM AGGREGC(INPUT,CUTPUT,TAPE1,TAPE2,TAPE3,TAPE4,
C
C *****MODIFIED 20/4/83 TO HANDLE MULTIPLE AREAS IN CLUSTER*****
C
C   +   TAPE5=INPUT,TAPE6=OUTPUT,TAPE7,TAPE9)
C
C PROGRAM ACCEPTS A LIST OF ENUMERATION DISTRICTS AS THE
C BASIS FOR CREATING A LARGER, USER-DEFINED AREA. FROM
C THIS LIST IT EXTRACTS THE OUTER BOUNDING SEGMENTS OF THE
C AGGREGATION AND WRITES A LIST OF THESE, IN A CLOCKWISE
C ORDER, TO A NEW FILE
C
C
C CALL SUBROUTINE TO EXTRACT SECTION NAMES FOR EACH E.D.
C FROM THE E.D. (POLYGON) FILE
C
C   CALL EDSEGN
C
C CALL SUBROUTINE TO FIND THE BOUNDARY OF EACH AREA
C
C   IFLG=0
C   CALL AREA3Y(IFLG)
C   IF (IFLG.NE.0) GOTO 9999
C
C CALL SUBROUTINE TO SORT THE FILE LISTING UNORDERED
C SEGMENT NAMES INTO A CLOCKWISE ORDER
C
C   IFLG=0
C   CALL BDISRT(IFLG)
C
C 9999  STOP
C      END
C
C
C
C   SUBROUTINE EDSEGN
C
C SUBROUTINE CREATES A FILE CONTAINING SEGMENT NAMES FOR
C ALL ENUMERATION DISTRICTS IN EACH AREA, USING EDFIL AND
C POLFIL AS INPUT, EDSECT IS CREATED
C
C   CHARACTER EDCCODE*4,AREANM*(3),FIRST3*3,FIRST1*1,SECT*8
C   REWIND 3
C
C READ THE AREA/E.D. FILE AND SEARCH THE POLYGON FILE FOR
C THE BOUNDING SEGMENTS (EDS MUST BE LISTED IN ASCENDING ORDER)
C
C 301  READ (3,3000,END=399,ERR=301)EDCCODE,AREANM
C 3000  FORMAT(A4,1X,A8)
C
C CHECK FOR START OF A NEW AREA, IF FOUND WRITE THE NAME
C TO A NEW FILE
C
C   IF (EDCCODE.NE.'AREA') GOTO 300
C   WRITE(4,3000) EDCCODE,AREANM
C   REWIND 1
C   GOTO 301
C
C IF AN E.D. CODE IS READ, LOOK FOR ITS SEGMENTS AND WRITE
C TO A NEW FILE
C
C 300  READ (1,3001,END=30,ERR=300)FIRST3

```

```

3001  FORMAT(5X,A8)
      IF (FIRST8.EQ.AREANM) GOTO 302
      GOTO 300
C
C IF END OF FILE IS FOUND, ASSUME THIS AREA IS COMPLETE
C
302  READ (1,3003,END=301) FIRST1,SECT
3003  FORMAT(A1,A8)
C
C WHEN THIS E.D. IS COMPLETE GO BACK FOR THE NEXT ONE
C
      IF (FIRST1.EQ.'A') THEN
          BACKSPACE 1
          GOTO 301
      ENDIF
      WRITE (4,3003) FIRST1,SECT
      GOTO 302
C
C
C WRITE ERROR MESSAGE TO OUTPUT
30  WRITE(6,3002) EDCCCE,AREANM
3002  FORMAT(1H0,' E.D. CODE ',A4,' IN AREA ',A8,
+ ' IS NOT FOUND ON FILE, SO IGNORED ')
      REWIND 1
      GOTO 301
C
399  RETURN
      END
C
C
C
      SUBROUTINE AREABY (IFLG)
C
C SUBROUTINE EXAMINES THE SECTION NAMES FOR EACH AREA.
C THOSE THAT ARE NOT MENTIONED TWICE WILL BE THE ONES
C THAT MAKE UP THE OUTER BOUNDARY OF THE AREA
C
      CHARACTER AREA*4, AREANM*(9),JOIN#1
      DIMENSION ITABLE (4,1500)
      LOGICAL ECF
      EOF=.FALSE.
      REWIND 4
C
C THE 4 CCLS OF THE ITABLE ARE AS FOLLOWS:
C 1=1ST NODE IN SEGMENT NAME
C 2=2ND NODE
C 3=REPRESENTATION OF CHARACTER IN BETWEEN
C 4=FLAG TO INDICATE WHETHER THIS SEGMENT HAS BEEN ACCESSED BEFORE
C
C
C SEGMENT DETAILS FOR THE FIRST (NEXT) AREA ARE READ FROM
C EDSECT AND PUT INTO A TABLE.  FIRSTLY, READ THE AREA NAME
C AND WRITE IT TO FILE
C
401  JROW = 0
      IF (EOF) GOTO 499
      READ(4,4000,END=499,ERR=413)AREA,AREANM
4000  FORMAT (A+,1X,A8)
      IF (AREA.NE.'AREA') GOTO 413
      WRITE (7,4000)AREA,AREANM
C
C NEXT READ THE SEGMENT NAMES IN THIS AREA AND

```

```

C PUT THEM INTO THE TABLE
C
412 READ (4,4002,END=415,ERR=400)NOD5A,JOIN,NODEB
4002 FORMAT (I4,A1,I4)
      JROW=JROW+1
      ITABLE(1,JROW)=NOD5A
      ITABLE(2,JROW)=NODEB
      IF (JOIN.EQ.'+') ITABLE(3,JROW)=0
      IF (JOIN.EQ.'A') ITABLE(3,JROW)=1
      IF (JOIN.EQ.'B') ITABLE(3,JROW)=2
      ITABLE(4,JROW)=0
      GOTO 412

C
C WHEN A NEW AREA IS FOUND, PROCESS THE ONE JUST COMPLETED
C JEND HOLDS THE NO. OF ENTRIES IN THE TABLE
C
400 BACKSPACE 4
414 JEND=JROW
407 JROW=0
403 IF (JROW.EQ.JEND) GOTO 406
      JROW=JROW+1

C
C TAKE EACH ROW OF THE TABLE IN TURN AND INSPECT THE REST OF
C THE TABLE FOR A MATCHING SECTION NAME, WHEN THE SEGMENT
C HAS BEEN MATCHED IT IS FLAGGED TO PREVENT IT BEING PROCESSED
C AGAIN
C
      IF (ITABLE(4,JROW).NE.0) GOTO 403
      NOD1=ITABLE(1,JROW)
      NOD2=ITABLE(2,JROW)
      JC = ITABLE(3,JROW)
      JPOSN = JROW
      IF (JPOSN.EQ.JEND) GOTO 405
408 JROW=JROW+1
      NNOD1=ITABLE(1,JROW)
      NNOD2=ITABLE(2,JROW)
      NJC = ITABLE(3,JROW)
      IF (NNOD1.EQ.NOD1.AND.NNOD2.EQ.NOD2.AND.NJC.EQ.JC) GOTO 404
      IF (JROW.EQ.JEND)GOTO 405
      GOTO 408

C
C WHEN TWO SEGMENTS HAVE BEEN FOUND TO MATCH (I.E. AT JROW AND
C JPOSN) THEY ARE FLAGGED WITH A 2, IF NO MATCH IS FOUND FOR A
C SEGMENT (AT JPOSN) IT IS FLAGGED WITH A 1.
C
404 ITABLE(4,JROW)=2
      ITABLE(4,JPOSN)=2
      GOTO 407
405 ITABLE(4,JPOSN)=1
      GOTO 407

C
C WHEN ALL THE TABLE ENTRIES HAVE BEEN FLAGGED (1=ONE SECTION
C WITH THIS NAME, 2=TWO SECTIONS) WRITE A LIST OF THE SECTION
C NAMES WITH FLAG=1 TO FILE
C
406 DO 42 JROW=1,JEND
      IF (ITABLE(4,JROW).EQ.2) GOTO 42
      IF (ITABLE(3,JROW).EQ.0) JCIN='+'
      IF (ITABLE(3,JROW).EQ.1) JCIN='A'
      IF (ITABLE(3,JROW).EQ.2) JCIN='B'
      ICHANL=7
      ISTNOC=ITABLE(1,JROW)

```



```

        IENNOD=ITABLE(2,JROW)
        CALL WRITEF(ISTNOD,JOIN,IENNOD,ICHANL)
42    CONTINUE
C
C DETAILS FOR THIS AREA ARE COMPLETE SO CLEAR ITABLE AND GO
C BACK FOR NEXT AREA
C
410    DO 40,J=1,300
        DO 41,I=1,4
            ITABLE(I,J) = 0
41    CONTINUE
40    CONTINUE
        GOTO 401
C
C SET EOF FLAG WHEN END OF FILE IS REACHED ON READING
C SEGMENT NAMES
C
415    EOF=.TRUE.
        GOTO 414
C
413    WRITE(6,4003)
4003    FORMAT(1H0," ERROR IN READING AREA NAME IN EDSECT.CAT ")
        IFLG=1
C
499    RETURN
        END
C
C
C
        SUBROUTINE BOYSRT (IFLG)
C
C SUBROUTINE SORTS THE FILE OF AREA BOUNDARY SEGMENT NAMES SO THAT
C THESE ARE LISTED IN CLOCKWISE ORDER
C
        LOGICAL CLOCK,EOF
        DIMENSION ITABLE(4,1500), ITABL2(3,1500)
        CHARACTER AREA*4,AREANM*(8),JOIN*1
        EOF=.FALSE.
        REWIND 7
C
C FIRSTLY, CLEAR THE TABLE, THEN READ THE SEGMENTS FOR THE FIRST
C (NEXT) AREA INTO IT. AREA NAME IS THE FIRST LINE READ
507    DO 50,J=1,1500
        DO 51,I=1,4
            ITABLE(I,J)=0
51    CONTINUE
50    CONTINUE
        IF (EOF) GOTO 599
        READ(7,5000,END=599,ERR=513)AREA,AREANM
5000    FORMAT(A4,1X,A8)
        IF (AREA.NE."AREA") GOTO 513
        JROW=0
C
C NEXT, READ ALL THE SEGMENTS FOR THIS AREA INTO ITABLE
C
501    READ(7,5001,END=509,ERR=500)NODCA,JOIN,NODEB
5001    FORMAT(I4,A1,I4)
        JROW=JROW+1
        ITABLE(1,JROW)=NODCA
        ITABLE(2,JROW)=NODEB
        IF (JOIN.EQ."+") ITABLE(3,JROW)=0
        IF (JOIN.EQ."A") ITABLE(3,JROW)=1

```

```

      IF (JOIN.EQ.'2') ITABLE (3,JROW)=2
      GOTO 501
C
C  WHEN A NEW AREA IS FOUND PROCESS THIS ONE
C
      500  BACKSPACE 7
      504  JEND = JROW
           JTEND = JROW+1
C
C (JROW IS USED HERE AS A COUNTER FOR ITABLE, JTAB FOR ITABL2)
C
      JROW=1
      JTAB=1
C
C THE FIRST SEGMENT IS USED AS THE STARTING PCINT, THE
C TABLE IS THEN READ TO FIND THE ADJOINING SEGMENT AND
C THIS IS WRITTEN TO A 2ND TABLE. THIS PROCESS IS REPEATED
C UNTIL THE PCLYGON IS COMPLETE.
C
      START = ITABLE(1,JROW)
      SNCDE = ITABLE(2,JROW)
      JC = ITABLE(3,JROW)
      ITABLE(4,JROW)=1
C
C WRITE THE FIRST SEGMENT IMMEDIATELY TO ITABL2
C
      ITABL2(1,JTAB)=START
      ITABL2(2,JTAB)=SNCDE
      ITABL2(3,JTAB)=JC
      55  JROW=JROW+1
           IF (JROW.EQ.JTEND) JROW=1
           IF (ITABLE(4,JROW).EQ.1) GOTO 55
      NODE1=ITABLE(1,JROW)
      NODE2=ITABLE(2,JROW)
      JC=ITABLE(3,JROW)
      IF (NODE1.EQ.SNCDE) GOTO 502
      IF (NODE2.EQ.SNCDE) GOTO 503
      GOTO 55
C
C PROCESSING IF THE 1ST NODE IN THE SEGMENT NAME READ IS A
C MATCH - I.E. THE SEGMENT NAME IS WRITTEN TO ITABL2 AND
C THEN A MATCH IS SOUGHT FOR THE 2ND NODE. IF IT MATCHES THE
C VERY FIRST NODE IN ITABL2 THEN THE PCLYGON IS COMPLETE
C
      502  JTAB=JTAB+1
           ITABL2(1,JTAB)=NODE1
           ITABL2(2,JTAB)=NODE2
           ITABL2(3,JTAB)=JC
           ITABLE(4,JROW)=1
           IF (NODE2.EQ.START) GOTO 506
           SNCDE = NODE2
           GOTO 55
C
C PROCESSING IF THE 2ND NODE IN THE SEGMENT NAME FOUND IS
C A MATCH
C
      503  JTAB=JTAB+1
           ITABL2(1,JTAB)=NODE1
           ITABL2(2,JTAB)=NODE2
           ITABL2(3,JTAB)=JC
           ITABLE(4,JROW)=1
           IF (NODE1.EQ.START) GOTO 506

```

```

        SNODE = NOD21
        GOTO 55
C
C CALCULATE NUMBER OF SEGMENTS WITHIN THIS AREA SO THAT
C CORRECT NUMBER ARE WRITTEN TO OUTPUT FILE
C
506      KCUNT = 0
        DO 57 K=1,JEND
          IF (ITABLE(4,K).EQ.1) KCUNT=KCUNT+1
57      CONTINUE
        PRINT #,"KCUNT=",KCUNT
C
C ITABL2 NOW CONTAINS THE SEGMENTS LISTED IN ORDER, BUT THEIR
C DIRECTION STILL HAS TO BE ESTABLISHED (I.E. CLOCKWISE OR
C ANTICLOCKWISE)
C
        CALL DIRCTN(CLOCK,ITABL2,KCUNT)
        IF (.NOT.CLOCK) GOTO 508
        WRITE (2,5000)AREA,AREANM
        DO 54 JTAB=1,KCUNT
          IF (ITABL2(3,JTAB).EQ.0) JOIN="+"
          IF (ITABL2(3,JTAB).EQ.1) JOIN="A"
          IF (ITABL2(3,JTAB).EQ.2) JOIN="B"
          CALL WRITEF ( ITABL2(1,JTAB),JOIN,ITABL2(2,JTAB),2)
54      CONTINUE
        GOTO 510
C
C IF THE SEGMENTS ARE ANTICLOCKWISE THEY NEED TO BE REVERSED
C BEFORE WRITING TO FILE
C
508      WRITE (2,5000)AREA,AREANM
        DO 52 JTAB=KCUNT,1,-1
          IF (ITABL2(3,JTAB).EQ.0) JOIN="+"
          IF (ITABL2(3,JTAB).EQ.1) JOIN="A"
          IF (ITABL2(3,JTAB).EQ.2) JOIN="B"
          CALL WRITEF ( ITABL2(1,JTAB),JOIN,ITABL2(2,JTAB),2)
52      CONTINUE
        GOTO 510
C
C PROCESSING FOR MORE THAN ONE CONTIGUOUS AREA IN A CLUSTER
C
C FIRSTLY, ELIMINATE SEGMENTS ALREADY USED AND CREATE NEW ITABLE
C
510      L=0
        DO 56 K=1,JEND
          IF (ITABLE(4,K).EQ.1) GOTO 56
          L=L+1
          ITABLE(1,L)=ITABLE(1,K)
          ITABLE(2,L)=ITABLE(2,K)
          ITABLE(3,L)=ITABLE(3,K)
          ITABLE(4,L)=0
56      CONTINUE
        JROW=L
C
C WHEN ALL THE AREAS IN THE CLUSTER HAVE BEEN PROCESSED GO ON
C TO THE NEXT CLUSTER
C
        IF (JROW.EQ.0) GOTO 507
        GOTO 504
C
C SET END OF FILE FLAG WHEN LAST SEGMENT HAS BEEN READ
C

```



```

509   EOF=.TRUE.
      GOTO 504
C
C  ERROR STATEMENT
C
513   WRITE(6,5002)
5002  FORMAT (1H0,' ERROR IN READING SECTION NAME IN ABBRY1.DAT ')
      IFLG=1
C
599   RETURN
      END
C
C
C
      SUBROUTINE DIRCTN(CLOCK,ITABL2,JENC)
C
C  SUBROUTINE EXAMINES SEGMENT NAMES AND THE THEIR COORDINATES
C  TO ESTABLISH WHETHER THE TABLE HOLDS THESE NAMES IN A
C  CLOCKWISE OR ANTICLOCKWISE DIRECTION
C
      DIMENSION ITABL2(3,1500),XARR1(300),YARR1(300),XARR2(2500),
+      YARR2(2500),XYZ(5000),ISGTAB(3,3000)
      CHARACTER IEOS*1,JCIN*1
      LOGICAL CLOCK,FIRST
      JROW=0
      LENAR2=0
C
C  THE FIRST TIME THIS SUBROUTINE IS CALLED SET UP A TABLE
C  HOLDING THE SEGMENT NAMES (IN THE SAME FORMAT AS ITABL2)
C  THIS WILL BE USED TO LOCATE SPECIFIC SEGMENTS IN THE
C  COORDINATES FILE
C
      IF (.NOT.FIRST) GOTO 608
      REWIND 9
      FIRST=.FALSE.
      I=0
C
C
C  EACH RECORD CONTAINS COORDINATES FOR ONE AREA
602   READ (9,END=609)NOC1,JCIN,NOC2
      I=I+1
      ISGTAB(1,I)=NOC1
      ISGTAB(2,I)=NOC2
      IF (JCIN.EC.'+') ISGTAB(3,I)=0
      IF (JCIN.EC.'A') ISGTAB(3,I)=1
      IF (JCIN.EC.'B') ISGTAB(3,I)=2
C
      GOTO 602
609   IEND=I
C
C  GET EACH SEGMENT NAME FROM THE TABLE AND CHECK THAT ITS
C  COORDINATES FOLLOW A NATURAL SEQUENCE FROM THE LAST SEGMENT
C  J IS INITIALISED EACH TIME A NEW SEGMENT IS
C  PROCESSED
C
608   REWIND 9
      LOCN=0
606   JROW=JROW+1
      J=0
      IF (JROW.GT.IEND) GOTO 601
C
C  USE THE NODE NUMBERS HELD IN THE LAST ELEMENT OF

```

```

C ITABL2 TO ESTABLISH WHICH OF THE NODES IN THE FIRST
C SEGMENT IS THE STARTING NODE
C
      IF (JROW.EQ.1) THEN
        IF (ITABL2(1,JROW).EQ.ITABL2(1,JEND).OR.
+         ITABL2(1,JROW).EQ.ITABL2(2,JEND)) THEN
          IFLAG=0
        ELSE
          IFLAG=1
        ENDIF
      ELSE
        IF (ITABL2(1,JROW).EQ.NODE8) THEN
          IFLAG=0
        ELSE
          IFLAG=1
        ENDIF
      ENDIF
C
C LOOK IN ISGTAB FOR THE SECTION NAME AND USE POINTER TO FIND
C THE COORDINATES
C
C
607      J=J+1
          IF (J.GT.IEND) GOTO 660
          IF (ITABL2(1,JROW).EQ.ISGTAB(1,J).AND.
+          ITABL2(2,JROW).EQ.ISGTAB(2,J).AND.
+          ITABL2(3,JROW).EQ.ISGTAB(3,J)) THEN
            GOTO 603
          ELSE
            GOTO 607
          ENDIF
C
C
C DECIDE WHETHER REQUIRED SEGMENT IS BEFORE OR AFTER
C CURRENT POSITION OF POINTER IN DIGITISER FILE
603      IF (J.GT.LOCN) THEN
          LINES=J-LOCN
        ELSE
          REWIND 9
          LINES=J
        ENDIF
        LOCN=J
C
C EXECUTE DUMMY READ OF DIGITISER FILE TO MOVE POINTER
C TO CORRECT RECORD
C
          DO 60 K=1,LINES-1
            READ (9)
          CONTINUE
60
C
C
C READ COORDINATES INTO X AND Y ARRAYS
C
604      READ (9) NOD1,JCIN,NOD2,LENAR1,
+          (XARR1(I),YARR1(I),I=1,LENAR1)
C
C WHEN ALL COORDINATES FOR THIS SECTION HAVE BEEN READ ADD THEM
C ONTO THE AREA ARRAY
C
605      CALL COORDS(XARR1,YARR1,XARR2,YARR2,IFLAG,LENAR1,LENAR2)
          IF (IFLAG.EQ.0) THEN
            NCDEA=ITABL2(1,JROW)

```

```

        NCDEB=ITABL2(2,JRCW)
    ELSE
        NODEA=ITABL2(2,JRCW)
        NCDEB=ITABL2(1,JRCW)
    ENDIF
    GOTO 606
C
C CONVERT THE TWO AREA ARRAYS INTO G.A.G. FORMAT
C
    601  J=0
        DO 6 I=1,LENAR2
            J=J+1
            XYZ(J)=XARR2(I)
            J=J+1
            XYZ(J)=YARR2(I)
        6  CONTINUE
C
C REPEAT THE FIRST PAIR OF COORDINATES TO MAKE SURE THE
C POLYGON JOINS UP
C
        J=J+1
        XYZ(J)=XARR2(1)
        J=J+1
        XYZ(J)=YARR2(1)
C
        ND=2
        NN=J
C
C CALL THE DIRECTION-FINDING SUBROUTINE
C
        CALL A1DIRN(XYZ,ND,NN,CLOCK)
        GOTO 699
C
C ERROR STATEMENTS
C
    650      WRITE (6,6004) ITABL2(1,JRCW),ITABL2(3,JRCW),ITABL2(2,JROW)
    6004      FORMAT(1H0,' SEGMENT ',I4,1X,I1,1X,I4,' NOT FOUND ')
        GOTO 606
C
    699  RETURN
        END
C
C
        SUBROUTINE COCRDS(XARR1,YARR1,XARR2,YARR2,IFLAG,LENAR1,LENAR2)
C
C SUBROUTINE ACCEPTS AN ARRAY OF COORDINATES DEFINING A SEGMENT
C IN A DIGITISER FILE AND IF NECESSARY REVERSES THEIR ORDER
C TO ENSURE THAT COORDINATES FOLLOW CONSECUTIVELY FROM ONE SEGMENT
C TO THE NEXT
C
        DIMENSION XARR1(300),YARR1(300),XARR2(2500),YARR2(2500)
        J=LENAR2
        IF (IFLAG.EQ.1) GOTO 100
C
C COORDINATES ADDED TO END OF AREA ARRAY IN ORIGINAL ORDER
C
        DO 9 I=1,LENAR1
            J=J+1
            XARR2(J)=XARR1(I)
            YARR2(J)=YARR1(I)
        9  CONTINUE
        GOTO 199

```



```

C
C COORDINATES REVERSED AS THEY ARE ADDED TO ARRAY
C
100 DO 8 I=LENAR1,1,-1
      J=J+1
      XARR2(J)=XARR1(I)
      YARR2(J)=YARR1(I)
      8 CONTINUE
199 LENAR2=J
      RETURN
      END

C
C
      SUBROUTINE ALCIRN(XYZ,ND,NN,CLOCK)
C
C SUBROUTINE USES THE G.A.G. ALGORITHM TO CHECK THE DIRECTION
C IN WHICH AN AREA IS SPECIFIED
C
C XYZ IS AN ARRAY OF X AND Y COORDINATES
C ND IS THE LENGTH OF A RECORD (I.E. 2 - X AND Y)
C NN IS THE LENGTH OF ARRAY XYZ
C CLOCK IS RETURNED INDICATING DIRECTION
C
      LOGICAL CLOCK
      DIMENSION XYZ(NN)
      A=0.0
      NV=ND+1
C
C CALCULATE SIGN OF POLYGON AREA (SUM OF ALL (X-LAST X)*(Y+LAST Y)
C
      DO 1 IA=NV,NN,ND
        A=A+(XYZ(IA)-XYZ(IA-ND))*(XYZ(IA+1)+XYZ(IA-ND+1))
      1 CONTINUE
C
C IF A IS POSITIVE THEN DIRECTION IS CLOCKWISE
C
      CLOCK=A.GT.0.0
      RETURN
C
      END
C
      SUBROUTINE WRITEF (ISTNOD,JOIN,IENNOD,ICHANL)
C
C ENSURES THAT LEADING SPACES IN THE SEGMENT NAMES ARE NOT
C WRITTEN TO THE OUTPUT FILE AS SPACES
C
      CHARACTER*1 JOIN,IGNE,IZERO
      IZERO='0'
C
C FIRSTLY ESTABLISH HOW MANY, AND WHERE, ARE THE LEADING ZEROS
C
      IF (ISTNOD.LT.10.AND.IENNOD.LT.10)GOTO 1001
      IF (ISTNOD.LT.10.AND.IENNOD.LT.100)GOTO 1002
      IF (ISTNOD.LT.10.AND.IENNOD.LT.1000)GOTO 1003
      IF (ISTNOD.LT.10.AND.IENNOD.GE.1000)GOTO 1004
C
      IF (ISTNOD.LT.100.AND.IENNOD.LT.10)GOTO 1005
      IF (ISTNOD.LT.100.AND.IENNOD.LT.100)GOTO 1006
      IF (ISTNOD.LT.100.AND.IENNOD.LT.1000)GOTO 1007
      IF (ISTNOD.LT.100.AND.IENNOD.GE.1000)GOTO 1008
C
      IF (ISTNOD.LT.1000.AND.IENNOD.LT.10)GOTO 1009

```

```

      IF (ISTNOD.LT.1000.AND.IENNOD.LT.100)GOTO 1010
      IF (ISTNOD.LT.1000.AND.IENNOD.LT.1000)GOTO 1011
      IF (ISTNOD.LT.1000.AND.IENNOD.GE.1000)GOTO 1012
C
      IF (ISTNOD.GE.1000.AND.IENNOD.LT.10)GOTO 1013
      IF (ISTNOD.GE.1000.AND.IENNOD.LT.100)GOTO 1014
      IF (ISTNOD.GE.1000.AND.IENNOD.LT.1000)GOTO 1015
      IF (ISTNOD.GE.1000.AND.IENNOD.GE.1000)GOTO 1016
C
C ADJUST THE "WRITE" STATEMENT SO ALL ZERCS ARE WRITTEN TO FILE
C
C FORMAT 000N+000N
C
1001  WRITE(ICHANL,5001)IZERO,IZERO,IZERO,ISTNOD,JOIN,
      +      IZERO,IZERO,IZERO,IENNOD
5001  FORMAT (3A1,I1,A1,3A1,I1)
      GOTO 1999
C
C FORMAT 000N+00NN
C
1002  WRITE (ICHANL,5002)IZERO,IZERO,IZERO,ISTNOD,JOIN,
      +      IZERO,IZERO,IENNOD
5002  FORMAT(3A1,I1,A1,2A1,I2)
      GOTO 1999
C
C FORMAT 000N+0NNN
C
1003  WRITE(ICHANL,5003)IZERO,IZERO,IZERO,ISTNOD,JOIN,
      +      IZERO,IENNOD
5003  FORMAT (3A1,I1,A1,A1,I3)
      GOTO 1999
C
C FORMAT 000N+NNNN
C
1004  WRITE(ICHANL,5004)IZERO,IZERO,IZERO,ISTNOD,JOIN,IENNOD
5004  FORMAT(3A1,I1,A1,I4)
      GOTO 1999
C
C FORMAT 00NN+000N
C
1005  WRITE(ICHANL,5005)IZERO,IZERO,ISTNOD,JOIN,
      +      IZERO,IZERO,IZERO,IENNOD
5005  FORMAT(2A1,I2,A1,3A1,I1)
      GOTO 1999
C
C FORMAT 00NN+00NN
C
1006  WRITE(ICHANL,5006)IZERO,IZERO,ISTNOD,JOIN,
      +      IZERO,IZERO,IENNOD
5006  FORMAT(2A1,I2,A1,2A1,I2)
      GOTO 1999
C
C FORMAT 00NN+0NNN
C
1007  WRITE(ICHANL,5007) IZERO,IZERO,ISTNOD,JOIN,IZERO,IENNOD
5007  FORMAT(2A1,I2,A1,A1,I3)
      GOTO 1999
C
C FORMAT 00NN+NNNN
C
1008  WRITE (ICHANL,5008)IZERO,IZERO,ISTNOD,JOIN,IENNOD
5008  FORMAT(2A1,I2,A1,I4)

```

```

      GOTO 1999
C
C  FORMAT ONNN+000N
C
1009  WRITE (ICHANL,5009) IZERO,ISTNOC,JOIN,
      +      IZERO,IZERO,IZERO,IENNOG
5009  FORMAT (A1,I3,A1,3A1,I1)
      GOTO 1999
C
C  FORMAT ONNN+00NN
C
1010  WRITE(ICHANL,5010) IZERO,ISTNOD,JOIN,IZERO,IZERC,IENNOD
5010  FORMAT(A1,I3,A1,2A1,I2)
      GOTO 1999
C
C  FORMAT ONNN+0NNN
C
1011  WRITE (ICHANL,5011) IZERO,ISTNOD,JOIN,IZERO,IENNOD
5011  FORMAT(A1,I3,A1,A1,I3)
      GOTO 1999
C
C  FORMAT ONNN+NNNN
C
1012  WRITE(ICHANL,5012) IZERO,ISTNOD,JOIN,IENNOD
5012  FORMAT (A1,I3,A1,I4)
      GOTO 1999
C
C  FORMAT NNNN+000N
C
1013  WRITE (ICHANL,5013) ISTNOD,JOIN,IZERO,IZERC,IZERO,IENNOG
5013  FORMAT(I4,A1,3A1,I1)
      GOTO 1999
C
C  FORMAT NNNN+00NN
C
1014  WRITE(ICHANL,5014) ISTNOD,JOIN,IZERC,IZERO,IENNOG
5014  FORMAT (I4,A1,2A1,I2)
      GOTO 1999
C
C  FORMAT NNNN+0NNN
C
1015  WRITE(ICHANL,5015) ISTNOD,JOIN,IZERO,IENNOD
5015  FORMAT (I4,A1,A1,I3)
      GOTO 1999
C
C  FORMAT NNNN+NNNN
C
1016  WRITE(ICHANL,5016) ISTNOD,JOIN,IENNOG
5016  FORMAT (I4,A1,I4)
C
1999  RETURN
      END
** ECR **
** ECF **

```


If A has a positive value then the direction is clockwise.

A2.2.3 (i) Program MPHOLE

(ii) Identifies new small areas within a cluster which have a 'hole' (or 'island') in them. i.e. EDs within their outer boundary which belong to another cluster.

(iii) Input files.

a) File of area names and their bounding segments as output from MPAGGR.

b) File of unformatted segment coordinates (see MPAGGR).

Output

a) No files are produced. Output is to the printer. A message is printed when one area is found to be within another.

(iv) Program description.

a) Main program.

Initialises tables and arrays containing cluster, area and segment names. Controls processing for one cluster at a time. Calls subroutine BOUND.

b) Subroutine BOUND.

For each area within the cluster, BOUND sets up an array containing the coordinates of its boundary. It then controls processing for testing whether subsequent areas lie within this. Calls SEGPTS, COORDS and SROUND.

c) Subroutine SEGPTS.

Extracts segment coordinates from an unformatted file, returning these in arrays.

d) Subroutine COORDS.

See program MPAGGR.

e) Subroutine SROUND.

Self-documenting, see program listing. See also program

Program MPHOLE

```

PROGRAM MPHOLE
C
C PROGRAM EXAMINES OUTPUT FROM MPAGGR TO IDENTIFY NEW AREAS FORMED BY
C CLUSTERING PROCESS WHICH HAVE A "HOLE" IN THEM (IE. AN AREA FROM ONE
C CLUSTER IS COMPLETELY SURROUNDED BY AN AREA FROM ANOTHER CLUSTER).
C
      DIMENSION ISEG1(3,999),ITABL2(3,400),ISEG1(3,200),IPTR(500)
      CHARACTER NAME*8
      CHARACTER JCIN*1,ICLARR(100)*8
      LOGICAL EOF
      EOF=.FALSE.

C
C INITIALISATION
C
C FIND OUT NUMBER OF STARTING CLUSTER AND SET NCLUS
C
      READ (1,1000,ERR=101) NAME,IDIGT1,ICIGT2
C
C IF NO ERROR THEN FIRST CLUSTER HAS TWO DIGIT NUMBER
C
      NCLUS=IDIGT1*10+IDIGT2-1
      GOTO 350
101    NCLUS=IDIGT1-1
350    BACKSPACE 1
300    IF (EOF) GOTO 9999
      K=0
      L=0
      IPTR(1)=1

C
C READ LIST OF AREAS AND SEGMENTS WITHIN THE 1ST (NEXT) CLUSTER - UP TO
C 29 CLUSTERS ARE ALLOWED BY THE PROGRAM. ONE CLUSTER IS PROCESSED AT A
C TIME
C
      NCLUS=NCLUS+1

C
C (AN ERROR WILL BE FOUND ON READ IF THE CLUSTER IS NUMBERED 1-9, THIS
C IS EXPECTED AND MAY BE IGNORED)
C
500    READ (1,1000,ERR=100) NAME,IDIGT1,ICIGT2
1000   FORMAT(5X,A8,T10,2I1)
100    IF (NCLUS.LE.9) THEN
          IF (IDIGT1.NE.NCLUS) THEN
              GOTO 700
          ELSE
              GOTO 200
          ENDIF
      ELSE
          IF (NCLUS.LE.19) THEN
              IF (IDIGT2.NE.NCLUS-10) THEN
                  GOTO 700
              ELSE
                  GOTO 200
              ENDIF
          ELSE
              IF (IDIGT2.NE.NCLUS-20) THEN
                  GOTO 700
              ELSE
                  GOTO 200
              ENDIF
          ENDIF
      ENDIF
C

```

```

C FOR EACH AREA IN THE CLUSTER PUT ITS SEGMENTS INTO A TABLE (ISEGTB).
C ICLARR HOLDS THE NAMES OF EVERY AREA IN THE CLUSTER.
C
200      K=K+1
        ICLARR(K)=NAME
C
C
C                                     (IPTR HOLDS STARTING POSITIONS
C                                     FOR SEGMENTS FOR EACH AREA)
C
        IPTR(K)=L+1
        DO 9 I=1,3000
C
C (IF ERROR IS FOUND ON READ THEN NEW AREA MUST HAVE BEEN STARTED SO
C GO ON TO PROCESS THIS ONE)
C
        READ (1,1001,ERR=400,END=600)NOC1,JOIN,NOC2
1001     FORMAT (I4,A1,I4)
        L=L+1
        ISEGTB(1,L)=NOC1
        ISEGTB(2,L)=NOC2
        IF(JOIN.EQ.'+')ISEGTB(3,L)=0
        IF(JOIN.EQ.'A')ISEGTB(3,L)=1
        IF(JOIN.EQ.'B')ISEGTB(3,L)=2
9        CONTINUE
400      BACKSPACE 1
        GOTO 500
C
C AT END OF FILE PUT POSITION INTO POINTER ARRAY TO ASSIST PROCESSING
C LATER ON (DO LOOP 8 REQUIRES THIS ELSE A SPURIOUS VALUE WILL BE GIVEN
C TO JEND)
C
600      IPTR(K+1)=L+1
C
C PROCESSING FOR EACH CLUSTER
C
C INITIALISATION
C
        EOF=.TRUE.
700      BACKSPACE 1
        KEND=K
C
C
C                                     (KEND HOLDS NO. OF AREAS
C                                     IN THIS CLUSTER, ALSO =
C                                     LENGTH OF ISEG1)
C
        K1=0
C
C CREATE AN ARRAY (ISEG1) CONTAINING THE FIRST SEGMENTS OF ALL THE AREAS
C IN THE CLUSTER. THIS WILL BE USED FOR EXTRACTING SAMPLE POINTS TO
C ESTABLISH WHETHER ONE AREA LIES WITHIN ANOTHER. THE ASSUMPTION IS MADE
C THAT IF ONE POINT (PAIR OF COORDINATES) LIES WITHIN AN AREA THEN THEY
C ALL WILL.
C
        DO 7 I=1,KEND
        IPT=IPTR(I)
        ISEG1(1,I)=ISEGTB(1,IPT)
        ISEG1(2,I)=ISEGTB(2,IPT)
        ISEG1(3,I)=ISEGTB(3,IPT)
7        CONTINUE
C
C PUT SEGMENTS FOR 1ST (NEXT) AREA INTO A SEPARATE ARRAY (ITABL2) - THESE
C WILL BE USED TO CREATE A STRING OF BOUNDARY COORDINATES FOR SUBSEQUENT
C TESTING FOR A HOLE

```



```

C      800      K1=K1+1
C              J=0
C
C
C              (K1 HOLDS THE NO. OF THE
C              AREA WITHIN THE CLUSTER)
C
C      IF (K1.GT.KEND) GOTO 300
C      DO 8 I=IPTR(K1),IPTR(K1+1)-1
C      J=J+1
C      ITABL2(1,J)=ISEGTB(1,I)
C      ITABL2(2,J)=ISEGTB(2,I)
C      ITABL2(3,J)=ISEGTB(3,I)
C
C      8      CONTINUE
C              JEND = IPTR(K1+1) - IPTR(K1)
C
C
C              (JEND HOLDS THE NO. OF
C              SEGMENTS IN THIS AREA,
C              ALSO = LENGTH OF ITABL2)
C
C      CALL SUBROUTINE TO PERFORM MAIN PROCESSING
C
C      PRINT *, '***** CALLING SUBROUTINE BOUND...'
C      PRINT *, 'JEND = ', JEND
C      PRINT *, 'KEND = ', KEND
C      PRINT *, 'K1 = ', K1
C      CALL BOUND(ISEG1,ITABL2,JEND,KEND,K1,ICLARR)
C      GOTO 800
C
C      9999     STOP
C              END
C
C
C      SUBROUTINE BOUND (ISEG1,ITABL2,JEND,KEND,K1,ICLARR)
C
C      SUBROUTINE SEARCHES AN UNFORMATTED FILE OF SEGMENT COORDINATES, CREATES
C      A CONTINUOUS BOUNDARY FOR THE AREA BEING CONSIDERED, AND LOOKS AT THE
C      FIRST POINT OF EVERY OTHER AREA IN THE CLUSTER IN ORDER TO ESTABLISH
C      WHETHER OR NOT ONE AREA LIES WITHIN ANOTHER. (IE. IF THERE IS AN AREA
C      WITH A HOLE IN IT
C
C      DIMENSION ITABL2(3,400),XARR1(300),YARR1(300),XARR2(3000),
C      + YARR2(3000),ISGTAB(3,3000),ISEG1(3,200),X(99),Y(99)
C      CHARACTER JOIN#1,ICLARR(100)*8
C      LOGICAL FIRST
C      JROW=0
C      LENAR2=0
C
C      THE FIRST TIME THIS SUBROUTINE IS CALLED SET UP A TABLE
C      HOLDING THE SEGMENT NAMES (IN THE SAME FORMAT AS ITABL2)
C      THIS WILL BE USED TO LOCATE SPECIFIC SEGMENTS IN THE
C      COORDINATES FILE
C
C      IF (.NOT.FIRST) GOTO 608
C      REWIND 9
C      FIRST=.FALSE.
C      I=0
C
C
C      EACH RECORD CONTAINS COORDINATES FOR ONE SEGMENT
C
C      602      READ (9,END=609)NCD1,JOIN,NCD2
C              I=I+1
C              ISGTAB(1,I)=NCD1

```

```

        ISGTAB(2,I)=NOD2
        IF (JOIN.EQ.'+') ISGTAB(3,I)=0
        IF (JOIN.EQ.'A') ISGTAB(3,I)=1
        IF (JOIN.EQ.'B') ISGTAB(3,I)=2
C
        GOTO 602
609      IEND=I
C
C GET EACH SEGMENT NAME FROM THE TABLE AND CHECK THAT ITS
C COORDINATES FOLLOW A NATURAL SEQUENCE FROM THE LAST SEGMENT
C J IS INITIALISED EACH TIME A NEW SEGMENT IS
C PROCESSED
C
608      REWIND 9
        LOCN=0
606      JROW=JROW+1
        IF (JROW.GT.IEND) GOTO 601
C
C USE THE NODE NUMBERS HELD IN THE LAST ELEMENT OF
C ITABL2 TO ESTABLISH WHICH OF THE NODS IN THE FIRST
C SEGMENT IS THE STARTING NODE
C
        IF (JROW.EQ.1) THEN
            IF (ITABL2(1,JROW).EQ.ITABL2(1,IEND).OR.
+           ITABL2(1,JROW).EQ.ITABL2(2,IEND)) THEN
                IFLAG=0
            ELSE
                IFLAG=1
            ENDIF
        ELSE
            IF (ITABL2(1,JROW).EQ.NOD2) THEN
                IFLAG=0
            ELSE
                IFLAG=1
            ENDIF
        ENDIF
C
C LOCATE THE SEGMENT IN THE UNFORMATTED FILE AND READ OFF ITS COORDINATES
C
        CALL SEGPTS(ITABL2,ISGTAB,JROW,LOCN,XARR1,YARR1,LENAR1)
C
C WHEN ALL COORDINATES FOR THIS SECTION HAVE BEEN READ ADD THEM
C ONTO THE AREA ARRAY
C
605      CALL COORCS(XARR1,YARR1,XARR2,YARR2,IFLAG,LENAR1,LENAR2)
        IF (IFLAG.EQ.0) THEN
            NOD2A=ITABL2(1,JROW)
            NOD2B=ITABL2(2,JROW)
        ELSE
            NOD2A=ITABL2(2,JROW)
            NOD2B=ITABL2(1,JROW)
        ENDIF
        GOTO 606
C
C WHEN A BOUNDARY HAS BEEN CREATED LOOK AT THE FIRST POINT OF EVERY OTHER
C AREA IN THE CLUSTER. FIRSTLY, SET UP AN ARRAY WITH THESE POINTS IN IT.
C
601      DO 61 JJ=1,KEND
        CALL SEGPTS(ISEG1,ISGTAB,JJ,LOCN,XARR1,YARR1,LENAR1)
        X(JJ)=XARR1(1)
        Y(JJ)=YARR1(1)
61      CONTINUE

```

```

C
C USE THESE POINTS FOR POINT-IN-POLYGON TEST. K1 IS THE SEQUENCE NUMBER
C OF THE AREA WHOSE BOUNDARY IS BEING TESTED, THEREFORE ITS FIRST POINT
C IS IGNORED (OTHERWISE IT WOULD BE FOUND TO BE LYING WITHIN ITSELF).
C
      DO 62 KK=1,KEND
      IF (KK.EQ.K1) GOTO 62
C
C SET INSIDE TO DUMMY VALUE
C
      INSIDE=9
      CALL SROUND(XARR2,YARR2,LENAR2,X(KK),Y(KK),INSIDE)
C
C IF THE POINT LIES ON THE BOUNDARY PRINT AN ERROR MESSAGE. (THIS
C SHOULD NEVER HAPPEN)
C
      IF (INSIDE.EQ.2.OR.INSIDE.EQ.3) PRINT *,
+      'ERROR : AREA ',ICLARR(KK),'SHARES BOUNDARY WITH AREA ',
+      ICLARR(K1)
C
C IF THE POINT LIES WITHIN THE BOUNDARY PRINT A WARNING MESSAGE SO THAT
C APPROPRIATE ACTION CAN BE TAKEN
C
      IF (INSIDE.EQ.1)PRINT *,
+      'AREA ',ICLARR(KK),KK,'LIES WITHIN AREA ',ICLARR(K1),K1
C
62      CONTINUE
699      RETURN
      END
C
      SUBROUTINE SEGPTSC(ITABL2,ISGTAB,JROW,LOCN,XARR1,YARR1,LENAR1)
C
C SUBROUTINE EXTRACTS SEGMENT COORDINATES FROM AN UNFORMATTED DIGITISER
C FILE
C
      CHARACTER JCIN#1
      DIMENSION ITABL2(3,400),ISGTAB(3,3000),XARR1(300),YARR1(300)
C
C LOCK IN ISGTAB FOR THE SECTION NAME AND USE PCINTER TO FIND
C THE COORDINATES
C
C
      J=0
607      J=J+1
      IF (ITABL2(1,JROW).EQ.ISGTAB(1,J).AND.
+      ITABL2(2,JROW).EQ.ISGTAB(2,J).AND.
+      ITABL2(3,JROW).EQ.ISGTAB(3,J)) THEN
          GOTO 603
      ELSE
          GOTO 607
      ENDIF
C
C
C DECIDE WHETHER REQUIRED SEGMENT IS BEFORE OR AFTER
C CURRENT POSITION OF POINTER IN DIGITISER FILE
603      IF (J.GT.LOCN) THEN
          LINES=J-LOCN
      ELSE
          REWIND 9
          LINES=J
      ENDIF
      LOCN=J

```



```

C
C EXECUTE DUMMY READ OF DIGITISER FILE TO MOVE POINTER
C TO CORRECT RECORD
C
      DO 60 K=1,LINES-1
      READ (9)
60    CONTINUE
C
C
C READ COORDINATES INTO X AND Y ARRAYS
C
      READ (9) NCD1,JCIN,NCD2,LENAR1,
+          (XARR1(I),YARR1(I),I=1,LENAR1)
C
      RETURN
      END
C
C
      SUBROUTINE COORDS(XARR1,YARR1,XARR2,YARR2,IFLAG,LENAR1,LENAR2)
C
C SUBROUTINE ACCEPTS AN ARRAY OF COORDINATES DEFINING A SEGMENT
C IN A DIGITISER FILE AND IF NECESSARY REVERSES THEIR ORDER
C TO ENSURE THAT COORDINATES FOLLOW CONSECUTIVELY FROM ONE SEGMENT
C TO THE NEXT
C
      DIMENSION XARR1(300),YARR1(300),XARR2(3000),YARR2(3000)
      J=LENAR2
      IF (IFLAG.EQ.1) GOTO 100
C
C COORDINATES ADDED TO END OF AREA ARRAY IN ORIGINAL ORDER
C
      DO 9 I=1,LENAR1
      J=J+1
      XARR2(J)=XARR1(I)
      YARR2(J)=YARR1(I)
9    CONTINUE
      GOTO 199
C
C COORDINATES REVERSED AS THEY ARE ADDED TO ARRAY
C
100  DO 8 I=LENAR1,1,-1
      J=J+1
      XARR2(J)=XARR1(I)
      YARR2(J)=YARR1(I)
8    CONTINUE
199  LENAR2=J
      RETURN
      END
      SUBROUTINE SROUND(XARRAY,YARRAY,NPOINT,XTEST,YTEST,INSIDE)
C
C A SUBROUTINE TO TEST WHETHER (XTEST,YTEST) IS INSIDE THE POLYGON
C DEFINED BY THE NPOINT VERTICES (XARRAY(I),YARRAY(I)).
C
C INSIDE RETURNS THE VALUE 0 IF THE POINT IS OUTSIDE THE POLYGON
C          1 IF IT IS INSIDE
C          2 IF IT IS ON A SIDE OF THE POLYGON
C          3 IF IT IS A VERTEX OF THE POLYGON
C
C THE TEST IS WHETHER AN ODD OR EVEN NUMBER OF SIDES OF THE POLYGON
C INTERSECT WITH A TEST LINE EXTENDING FROM THE TEST POINT VERTICALLY
C UPWARDS. THE LOGICAL VARIABLE IN IS INITIALLY FALSE AND IS
C NEGATED WHENEVER AN INTERSECTION IS FOUND, SO IF IT ENDS UP TRUE

```

```

C THE THE TEST POINT IS INSIDE THE POLYGON.
C
C IF THE TEST POINT IS FOUND TO BE ON A SIDE OR VERTEX, THERE IS
C NO NEED TO CONTINUE EXAMINING FURTHER SIDES OF THE POLYGON.
C
C A SIDE MEETS THE TEST LINE IF IT X-COORDINATES ARE ON EITHER SIDE
C OF XTEST AND ITS Y-COORDINATE AT THE POSITION XTEST IS GREATER THAN
C YTEST.
C
C AWKWARD CASES ARISE IF XTEST IS EQUAL AN X-COORDINATE OF A VERTEX.
C THIS CASE IS CONSIDERED ONLY WHEN THE X-COORDINATE IS THE SECOND
C ONE FOR THE SIDE: THIS ENSURES THAT THE CASE IS DEALT WITH ONLY ONCE
C FOR THAT SIDE.
C
C IF THE Y-COORDINATE IS LESS THAN YTEST THEN THERE IS NO INTERSECTION.
C IF THE Y-COORDINATE IS EQUAL TO YTEST THEN THE TEST POINT COINCIDES
C WITH A VERTEX.
C IF THE Y-COORDINATE IS GREATER THAN YTEST, THEN THE TEST LINE PASSES
C THROUGH A VERTEX.
C
C IN THIS CASE THERE IS AN INTERSECTION
C IF THE NEXT SIDE TO BE CONSIDERED LIES ON THE OPPOSITE SIDE OF THE TEST
C LINE FROM THE CURRENT SIDE. THAT WILL BE SO IF XTEST IS BETWEEN THE
C FIRST X-COORDINATE OF THE CURRENT SIDE AND THE NEXT X-COORDINATE TO BE
C CONSIDERED.
C IF THE NEXT X-COORDINATE LIES ON THE SAME SIDE OF THE TEST LINE AS THE
C CURRENT FIRST X-COORDINATE, THEN THERE IS NO INTERSECTION.
C IF THE NEXT X-COORDINATE IS EQUAL TO XTEST, THEN THE TEST POINT LIES
C ON THE NEXT SIDE, POSSIBLY PROJECTED, I.E. THE TEST LINE IS COLINEAR
C WITH A SIDE.
C
C IF THE Y-COORDINATE IS LESS THAN YTEST, THEN THE TEST POINT LIES ON A
C SIDE OF THE POLYGON.
C IF THE Y-COORDINATE IS EQUAL TO YTEST, THEN THE TEST POINT COINCIDES WITH
C A VERTEX OF THE POLYGON.
C IF THE Y-COORDINATE IS GREATER THAN YTEST, IT IS NECESSARY TO LOOK AHEAD
C UNTIL AN X-COORDINATE IS FOUND WHICH IS NOT COLINEAR WITH THE TEST LINE.
C THEN IF THAT X-COORDINATE IS ON THE OPPOSITE SIDE OF THE TEST LINE TO
C THE CURRENT FIRST X-COORDINATE, THERE IS AN INTERSECTION. IF IT IS ON
C THE SAME SIDE, THERE IS NO INTERSECTION. AFTER LOOKING AHEAD FOR SO
C MANY SIDES, THE NORMAL TESTING IS RESUMED WITH THE NEXT SIDE AFTER THAT
C JUST LOOKED AT.
C
C SINCE THE FIRST SIDE OF THE POLYGON TO HAVE BEEN TESTED MAY HAVE BEEN
C WITHIN A SEQUENCE OF SIDES COLINEAR WITH THE TEST LINE, IN LOOKING
C AHEAD IT MAY BE NECESSARY TO CONTINUE NOT JUST TO THE LAST SIDE BUT
C TO LOOK AGAIN AT ALL THE SIDES FROM THE FIRST UP TO THE CURRENT ONE.
C IF LOOKING AHEAD DOES NOT RESOLVE THE QUESTION OF WHETHER THERE IS AN
C INTERSECTION, THEN WE HAVE THE CASE WHERE ALL SIDES ARE COLINEAR WITH THE
C TEST LINE AND DO NOT MEET THE TEST POINT. I.E. THE "POLYGON" IS A
C VERTICAL LINE, AND HENCE THE TEST POINT IS NOT INSIDE IT.
C
      REAL XARRAY(NPCINT), YARRAY(NPOINT)
      LOGICAL IN
C
C UNTIL AN INTERSECTION IS FOUND ASSUME THE TEST POINT IS NOT INSIDE
C THE POLYGON.
C
      IN = .FALSE.
C
C GO THROUGH ALL VERTICES OF THE POLYGON
C

```

```

      NP = 1
20    X1 = XARRAY(NP)
      Y1 = YARRAY(NP)
      X2 = XARRAY(MOD(NP,NPOINT)+1)
      Y2 = YARRAY(MOD(NP,NPOINT)+1)
C
C TEST WHETHER XTEST LIES BETWEEN X1 AND X2
C
      BETWEN = (XTEST-X1)*(XTEST-X2)
C
C IF IT DOES NOT, THEN GO ON TO THE NEXT SIDE OF THE POLYGON
C
      IF(BETWEN.GT.0) GOTO 10
      IF(BETWEN.EQ.0) GOTO 40
C
C IN THIS CASE VERTICES ARE ON EITHER SIDE OF THE TEST LINE.
C SEE WHETHER THE POINT OF INTERSECTION OF THE SIDE WITH THE
C TEST LINE IS ABOVE OR BELOW THE TEST POINT.
C
      YINT = Y1 + (XTEST-X1)*(Y2-Y1)/(X2-X1)
C
C IF THE INTERSECTION IS BELOW THE TEST POINT, THEN GO ON
C TO THE NEXT SIDE OF THE POLYGON
C
      IF(YINT.LT.YTEST) GOTO 10
      IF(YINT.EQ.YTEST) GOTO 50
C
C THE SIDE MEETS THE TEST LINE ABOVE THE TEST POINT:
C COUNT THIS INTERSECTION.
C
      IN = .NOT.IN
      GOTO 10
C
C THE MEETING POINT COINCIDES WITH THE TEST POINT, SO THE
C TEST POINT IS ON A SIDE OF THE POLYGON
C
50    INSIDE = 2
      RETURN
C
C THE TEST LINE PASSES THROUGH A VERTEX OF THE POLYGON.
C IF THE VERTEX IS THE FIRST ONE OF THE CURRENT SIDE,
C THE PROBLEM HAS ALREADY BEEN DEALT WITH (OR WILL BE
C EVENTUALLY IF IT IS THE FIRST VERTEX OF THE POLYGON)
C AND SO WE GO ON TO CONSIDER THE NEXT VERTEX.
C
40    IF(XTEST.NE.X2) GOTO 10
C
C IF THE VERTEX IS BELOW THE TEST POINT, THERE IS NO
C INTERSECTION AND WE GO ON.
C
      IF(Y2.LT.YTEST) GOTO 10
      IF(Y2.NE.YTEST) GOTO 70
C
C THE TEST POINT COINCIDES WITH A VERTEX
C
      INSIDE = 3
      RETURN
C
C THE TEST LINE PASSES THROUGH A VERTEX. WE NOW LOOK
C AHEAD TO SUBSEQUENT SIDES TO SEE IF THIS IS TO COUNT
C AS AN INTERSECTION.
C

```



```

70  DO 110 NPX = NP+2,NP+NPOINT
    X3 = XARRAY(MOD(NPX-1,NPOINT)+1)
    BETWEN = (XTEST-X1)*(XTEST-X3)
C
C IF X1 AND X3 LIE ON THE SAME SIDE OF XTEST, THERE
C IS NO INTERSECTION
C
    IF(BETWEN.GT.0) GOTO 10
    IF(BETWEN.EQ.0) GOTO 80
C
C XTEST LIES BETWEEN X1 AND X3 SO THERE IS AN INTERSECTION
C
    IN = .NOT.IN
C
C THE SIDES LOOKED AT NEED NOT BE EXAMINED AGAIN
C
    NP = NPX
    GOTO 10
C
C IN THIS CASE THE TEST POINT IS COLINEAR WITH A SIDE OF
C THE POLYGON
C
80  Y3 = YARRAY(MOD(NPX-1,NPOINT)+1)
    IF(Y3.GE.YTEST) GOTO 90
C
C THE SIDE EXTENDS BELOW THE TEST POINT, AND HENCE
C PASSES VERTICALLY THROUGH IT.
C
    INSIDE = 2
    RETURN
90  IF(Y3.LE.YTEST) GOTO 110
C
C THE TEST POINT COINCIDES WITH A VERTEX
C
    INSIDE = 3
    RETURN
C
C THE TEST LINE IS COLINEAR WITH A SIDE. IT IS NOT YET
C DETERMINED WHETHER THIS COUNTS AS AN INTERSECTION, SO
C CONTINUE LOOKING AHEAD
C
110  CONTINUE
C
C IF WE HAVE LOOKED AHEAD THROUGH ALL VERTICES AGAIN AND
C NOT SETTLED WHETHER THERE IS AN INTERSECTION, THEN ALL
C SIDES MUST BE COLINEAR WITH THE TEST LINE BUT NONE PASS
C THROUGH THE TEST POINT. THEREFORE THE TEST POINT IS
C OUTSIDE THE LINEAR "POLYGON".
C
    INSIDE = 0
    RETURN
C
C END OF LOOP. GO TO NEXT VERTEX, IF ANY.
C
10  NP = NP+1
    IF(NP.LE.NPOINT) GOTO 20
C
C THE TEST POINT DOES NOT LIE ON A SIDE OR VERTEX OF THE
C POLYGON. IN HAS RECORDED WHETHER THE TEST POINT IS INSIDE
C OR OUTSIDE THE POLYGON.
C
    IF(IN) INSIDE = 1

```

```
IF(.NOT.IN) INSIDE = 0  
RETURN  
END
```

MPPTPL.

A2.2.4 (i) Program MPPTPL

(ii) Allocates postcodes to enumeration districts using a point-in-polygon routine.

(iii) Input files.

- a) Contains centroid coordinates of all postcodes.
- b) Contains strings of coordinates defining the bounding segments of all EDs.

Output files.

- a) Contains a list of postcodes and the EDs within which they lie.
- b) Contains a list of postcodes which lie on ED boundaries, and their respective EDs.
- c) Postcodes which do not lie within any area are output to the printer.

(iv) Program description.

a) Main program.

The centroid coordinates of each postcode are read in turn and for each one processing is as follows: coordinates are read for each ED and, having ensured that these are listed consecutively, the point-in-polygon routine is called. If the postcode is found to lie within this ED then both are written to file, failing this the next ED is tested. If the postcode is found to lie on a boundary it is written to a separate file and the search continues for the adjoining ED. The program ends when all postcodes have been processed. Calls subroutines COORDS and SROUND.

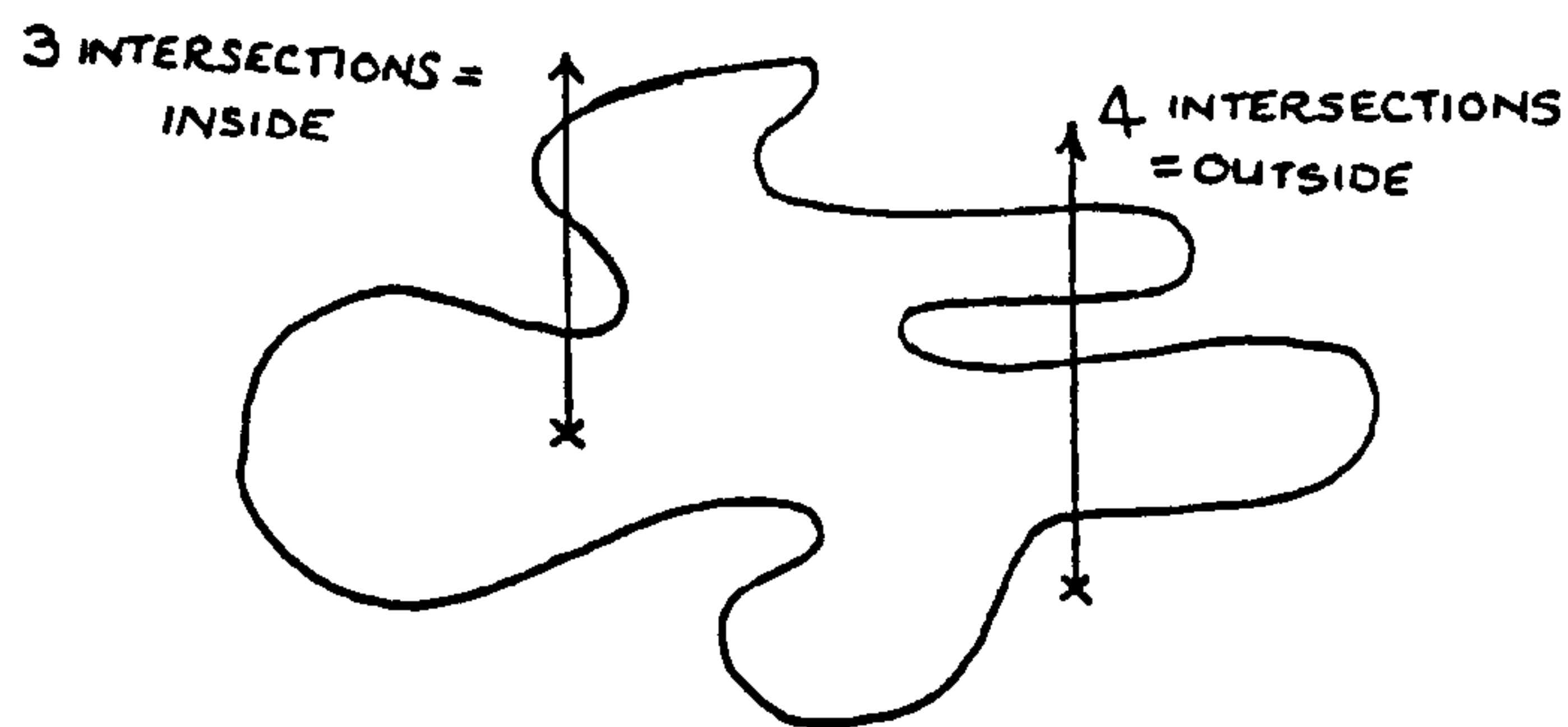
b) Subroutine COORDS.

See program MPAGGR.

- c) Subroutine SROUND (written by Dr. A. Matthews, Leicester University Computer Centre). Accepts arrays of x and y coordinates defining an area (ED) boundary and the coordinates of a point (postcode) and returns a flag showing whether the point lies within, outside or on the boundary of the area (see Notes below).

(v) Notes.

- a) The point-in-polygon algorithm used is the 'line-crossing method' which is explained in detail in the code. Simply, the point lies within the polygon if a vertical line drawn upwards from the point intersects with the boundary of the polygon an odd number of times



- b) In a second version of this program the subroutine SROUND was replaced by an alternative point-in-polygon routine. On a small test data set no improvement in processing time was achieved.
- c) A third version of the program utilised an unformatted ED boundary file, reducing the number of records to one per area. This reduced processing time to about one tenth of its previous level, now approximately 5 seconds per postcode on the full data set.
- d) Finally it was decided to adopt a two stage approach to allocating postcodes to EDs. Programs MPPTP4 and MPPTP5 were produced.

Program MPPTPL

```

PROGRAM MPPTPL (INPUT,OUTPUT,TAPE1,TAPE2,TAPE3,TAPE4,
+   TAPES=OUTPUT)
C
C PROGRAM TO TEST WHICH ENUMERATION DISTRICTS A SET OF
C POSTCODES LIE IN
C
    CHARACTER*4 PCODE,EDCODE*8,IFIRST*1,IAREA
    DIMENSION XARR1(100),YARR1(100),XARR2(500),YARR2(500)
    LOGICAL EOF,BORDER,FIRST
C
C READ POSTCODE FILE THEN SEARCH THROUGH E.D. FILE TO
C FIND WHICH E.D. THE POSTCODE LIES IN. IF IT LIES ON
C A BOUNDARY ESTABLISH WHICH E.D.S LIE ON EITHER SIDE
C
    106     EOF=.FALSE.
           BORDER=.FALSE.
           READ (1,1000,END=9999) XPC,YPC,PCODE
    1000    FORMAT (2F6.3,A4)
           REWIND 2
    101     IF (EOF) GOTO 107
           READ (2,1001,END=107) IAREA,EDCODE
    1001    FORMAT (A4,1X,A8)
           LENAR2=0
           FIRST=.TRUE.
C
C LOOK FOR THE START OF THE FIRST E.D.
C
           IF (IAREA.EQ.'AREA') GOTO 100
           GOTO 101
C
C READ THE RECORD CONTAINING THE SEGMENT NAME
C
    100     READ (2,1002,ERR=102,END=110) NCODE1,NCODE2
    1002    FORMAT (I4,1X,I4)
           I=0
           IF (FIRST) THEN
               IFLAG=0
               FIRST=.FALSE.
               GOTO 103
           ENDIF
           IF (NCODE1.EQ.NCODE2) THEN
               IFLAG=0
           ELSE
               IFLAG=1
           ENDIF
C
C READ THE COORDINATES OF THIS E.D. (POLYGON)
C
    103     READ (2,1003,END=112) IFIRST,XED,YED
    1003    FORMAT (A1,4X,E13.7,2X,E13.7)
C
           IF (IFIRST.EQ.'-') GOTO 111
C
C WRITE THE COORDINATES INTO THE TWO ARRAYS
           I=I+1
           XARR1(I)=XED
           YARR1(I)=YED
           GOTO 103
    111     LENAR1=I
           CALL COORDS(XARR1,YARR1,XARR2,YARR2,IFLAG,LENAR1,LENAR2)
           READ (2,1006)
    1006    FORMAT (1X)

```

```

      NODEA=NODE1
      NODEB=NODE2
      GOTO 100
102  BACKSPACE 2
109  NPTS=LENAR2
      INSIDE=9
C
C CALL THE POINT-IN-POLYGON ROUTINE
C
      CALL SROUND(XARR2,YARR2,NPTS,XPC,YPC,INSIDE)
C
C TEST FOR WHERE THE POSTCODE LIES IN RELATION TO THE E.D.
C
      IF (INSIDE.EQ.0) GOTO 101
      IF (INSIDE.EQ.1) GOTO 104
      IF (INSIDE.EQ.2.OR.INSIDE.EQ.3) GOTO 105
C
C IF POSTCODE IS FOUND TO BE INSIDE AN E.C. WRITE BOTH TO FILE
C
104  WRITE (3,1004) PCCDE,EDCCDE
1004  FORMAT (A4,1X,A3)
      GOTO 106
C
C SET FLAG WHEN END OF FILE HAS BEEN REACHED
C
110  EOF=.TRUE.
      GOTO 109
C
C IF POSTCODE IS ON THE BORDER OF AN E.D. WRITE BOTH TO AN
C ANCILLARY FILE AND KEEP LOOKING TO FIND OUT WHICH BORDER
C IT IS ON
105  WRITE (4,1004) PCCDE,EDCCDE
      BORDER=.TRUE.
      IF (EOF) GOTO 106
      GOTO 101
C
C IF END OF E.D. FILE IS FOUND THEN POSTCODE HAS NOT BEEN
C ALLOCATED SO DISPLAY AN ERROR MESSAGE
C
107  IF (BORDER) GOTO 106
      WRITE (5,1005) PCCDE
1005  FORMAT (1H0,' POSTCODE ',A4,' IS NOT IN ANY AREA')
      GOTO 106
C
112  WRITE(5,1007) NODE1, NODE2
1007  FORMAT(1H0,' EOF FOUND IN MIDDLE OF SECTION ',I4,1X,I4)
C
9999  STOP
      END
C
      SUBROUTINE SROUND(XARRAY,YARRAY,NPOINT,XTEST,YTEST,INSIDE)
C
C A SUBROUTINE TO TEST WHETHER (XTEST,YTEST) IS INSIDE THE POLYGON
C DEFINED BY THE NPOINT VERTICES (XARRAY(I),YARRAY(I)).
C
C INSIDE RETURNS THE VALUE 0 IF THE POINT IS OUTSIDE THE POLYGON
C 1 IF IT IS INSIDE
C 2 IF IT IS ON A SIDE OF THE POLYGON
C 3 IF IT IS A VERTEX OF THE POLYGON
C
C THE TEST IS WHETHER AN ODD OR EVEN NUMBER OF SIDES OF THE POLYGON
C INTERSECT WITH A TEST LINE EXTENDING FROM THE TEST POINT VERTICALLY

```



```

C UPWARDS. THE LOGICAL VARIABLE IN IS INITIALLY FALSE AND IS
C NEGATED WHENEVER AN INTERSECTION IS FOUND. SO IF IT ENDS UP TRUE
C THE THE TEST POINT IS INSIDE THE POLYGON.
C
C IF THE TEST POINT IS FOUND TO BE ON A SIDE OR VERTEX, THERE IS
C NO NEED TO CONTINUE EXAMINING FURTHER SIDES OF THE POLYGON.
C
C A SIDE MEETS THE TEST LINE IF IT X-COORDINATES ARE ON EITHER SIDE
C OF XTEST AND ITS Y-COORDINATE AT THE POSITION XTEST IS GREATER THAN
C YTEST.
C
C AWKWARD CASES ARISE IF XTEST IS EQUAL AN X-COORDINATE OF A VERTEX.
C THIS CASE IS CONSIDERED ONLY WHEN THE X-COORDINATE IS THE SECOND
C ONE FOR THE SIDE: THIS ENSURES THAT THE CASE IS DEALT WITH ONLY ONCE
C FOR THAT SIDE.
C
C IF THE Y-COORDINATE IS LESS THAN YTEST THEN THERE IS NO INTERSECTION.
C IF THE Y-COORDINATE IS EQUAL TO YTEST THEN THE TEST POINT COINCIDES
C WITH A VERTEX.
C IF THE Y-COORDINATE IS GREATER THAN YTEST, THEN THE TEST LINE PASSES
C THROUGH A VERTEX.
C
C IN THIS CASE THERE IS AN INTERSECTION
C IF THE NEXT SIDE TO BE CONSIDERED LIES ON THE OPPOSITE SIDE OF THE TEST
C LINE FROM THE CURRENT SIDE. THAT WILL BE SO IF XTEST IS BETWEEN THE
C FIRST X-COORDINATE OF THE CURRENT SIDE AND THE NEXT X-COORDINATE TO BE
C CONSIDERED.
C IF THE NEXT X-COORDINATE LIES ON THE SAME SIDE OF THE TEST LINE AS THE
C CURRENT FIRST X-COORDINATE, THEN THERE IS NO INTERSECTION.
C IF THE NEXT X-COORDINATE IS EQUAL TO XTEST, THEN THE TEST POINT LIES
C ON THE NEXT SIDE, POSSIBLY PROJECTED, I.E. THE TEST LINE IS COLINEAR
C WITH A SIDE.
C
C IF THE Y-COORDINATE IS LESS THAN YTEST, THEN THE TEST POINT LIES ON A
C SIDE OF THE POLYGON.
C IF THE Y-COORDINATE IS EQUAL TO YTEST, THEN THE TEST POINT COINCIDES WITH
C A VERTEX OF THE POLYGON.
C IF THE Y-COORDINATE IS GREATER THAN YTEST, IT IS NECESSARY TO LOOK AHEAD
C UNTIL AN X-COORDINATE IS FOUND WHICH IS NOT COLINEAR WITH THE TEST LINE.
C THEN IF THAT X-COORDINATE IS ON THE OPPOSITE SIDE OF THE TEST LINE TO
C THE CURRENT FIRST X-COORDINATE, THERE IS AN INTERSECTION. IF IT IS ON
C THE SAME SIDE, THERE IS NO INTERSECTION. AFTER LOOKING AHEAD FOR SO
C MANY SIDES, THE NORMAL TESTING IS RESUMED WITH THE NEXT SIDE AFTER THAT
C JUST LOOKED AT.
C
C SINCE THE FIRST SIDE OF THE POLYGON TO HAVE BEEN TESTED MAY HAVE BEEN
C WITHIN A SEQUENCE OF SIDES COLINEAR WITH THE TEST LINE, IN LOOKING
C AHEAD IT MAY BE NECESSARY TO CONTINUE NOT JUST TO THE LAST SIDE BUT
C TO LOOK AGAIN AT ALL THE SIDES FROM THE FIRST UP TO THE CURRENT ONE.
C IF LOOKING AHEAD DOES NOT RESOLVE THE QUESTION OF WHETHER THERE IS AN
C INTERSECTION, THEN WE HAVE THE CASE WHERE ALL SIDES ARE COLINEAR WITH THE
C TEST LINE AND DO NOT MEET THE TEST POINT. I.E. THE "POLYGON" IS A
C VERTICAL LINE, AND HENCE THE TEST POINT IS NOT INSIDE IT.
C
      REAL XARRAY(NPOINT), YARRAY(NPOINT)
      LOGICAL IN
C
C UNTIL AN INTERSECTION IS FOUND ASSUME THE TEST POINT IS NOT INSIDE
C THE POLYGON.
C
      IN = .FALSE.
C

```

```

C GO THROUGH ALL VERTICES OF THE POLYGON
C
  NP = 1
20  X1 = XARRAY(NP)
    Y1 = YARRAY(NP)
    X2 = XARRAY(MOD(NP,NPOINT)+1)
    Y2 = YARRAY(MOD(NP,NPOINT)+1)
C
C TEST WHETHER XTEST LIES BETWEEN X1 AND X2
C
  BETWEN = (XTEST-X1)*(XTEST-X2)
C
C IF IT DOES NOT, THEN GO ON TO THE NEXT SIDE OF THE POLYGON
C
  IF(BETWEN.GT.0) GOTO 10
  IF(BETWEN.EQ.0) GOTO 40
C
C IN THIS CASE VERTICES ARE ON EITHER SIDE OF THE TEST LINE.
C SEE WHETHER THE POINT OF INTERSECTION OF THE SIDE WITH THE
C TEST LINE IS ABOVE OR BELOW THE TEST POINT.
C
  YINT = Y1 + (XTEST-X1)*(Y2-Y1)/(X2-X1)
C
C IF THE INTERSECTION IS BELOW THE TEST POINT, THEN GO ON
C TO THE NEXT SIDE OF THE POLYGON
C
  IF(YINT.LT.YTEST) GOTO 10
  IF(YINT.EQ.YTEST) GOTO 50
C
C THE SIDE MEETS THE TEST LINE ABOVE THE TEST POINT:
C COUNT THIS INTERSECTION.
C
  IN = .NOT.IN
  GOTO 10
C
C THE MEETING POINT COINCIDES WITH THE TEST POINT, SO THE
C TEST POINT IS ON A SIDE OF THE POLYGON
C
50  INSIDE = 2
    RETURN
C
C THE TEST LINE PASSES THROUGH A VERTEX OF THE POLYGON.
C IF THE VERTEX IS THE FIRST ONE OF THE CURRENT SIDE,
C THE PROBLEM HAS ALREADY BEEN DEALT WITH (OR WILL BE
C EVENTUALLY IF IT IS THE FIRST VERTEX OF THE POLYGON)
C AND SO WE GO ON TO CONSIDER THE NEXT VERTEX.
C
40  IF(XTEST.NE.X2) GOTO 10
C
C IF THE VERTEX IS BELOW THE TEST POINT, THERE IS NO
C INTERSECTION AND WE GO ON.
C
  IF(Y2.LT.YTEST) GOTO 10
  IF(Y2.NE.YTEST) GOTO 70
C
C THE TEST POINT COINCIDES WITH A VERTEX
C
  INSIDE = 3
  RETURN
C
C THE TEST LINE PASSES THROUGH A VERTEX. WE NOW LOOK
C AHEAD TO SUBSEQUENT SIDES TO SEE IF THIS IS TO COUNT

```

```

C AS AN INTERSECTION.
C
70  DO 110 NPX = NP+2,NP+NPCINT
    X3 = XARRAY(MOD(NPX-1,NPOINT)+1)
    BETWEN = (XTEST-X1)*(XTEST-X3)
C
C IF X1 AND X3 LIE ON THE SAME SIDE OF XTEST, THERE
C IS NO INTERSECTION
C
    IF(BETWEN.GT.0) GOTO 10
    IF(BETWEN.EQ.0) GOTO 80
C
C XTEST LIES BETWEEN X1 AND X3 SO THERE IS AN INTERSECTION
C
    IN = .NOT.IN
C
C THE SIDES LOOKED AT NEED NOT BE EXAMINED AGAIN
C
    NP = NPX
    GOTO 10
C
C IN THIS CASE THE TEST POINT IS COLINEAR WITH A SIDE OF
C THE POLYGON
C
80  Y3 = YARRAY(MOD(NPX-1,NPOINT)+1)
    IF(Y3.GE.YTEST) GOTO 90
C
C THE SIDE EXTENDS BELOW THE TEST POINT, AND FENCE
C PASSES VERTICALLY THROUGH IT.
C
    INSIDE = 2
    RETURN
90  IF(Y3.LE.YTEST) GOTO 110
C
C THE TEST POINT COINCIDES WITH A VERTEX
C
    INSIDE = 3
    RETURN
C
C THE TEST LINE IS COLLINEAR WITH A SIDE. IT IS NOT YET
C DETERMINED WHETHER THIS COUNTS AS AN INTERSECTION, SO
C CONTINUE LOOKING AHEAD
C
110  CONTINUE
C
C IF WE HAVE LOOKED AHEAD THROUGH ALL VERTICES AGAIN AND
C NOT SETTLED WHETHER THERE IS AN INTERSECTION, THEN ALL
C SIDES MUST BE COLLINEAR WITH THE TEST LINE BUT NONE PASS
C THROUGH THE TEST POINT. THEREFORE THE TEST POINT IS
C OUTSIDE THE LINEAR "POLYGON".
C
    INSIDE = 0
    RETURN
C
C END OF LOOP. GO TO NEXT VERTEX, IF ANY.
C
10  NP = NP+1
    IF(NP.LE.NPOINT) GOTO 20
C
C THE TEST POINT DOES NOT LIE ON A SIDE OR VERTEX OF THE
C POLYGON. IN HAS RECORDED WHETHER THE TEST POINT IS INSIDE
C OR OUTSIDE THE POLYGON.

```



```

C      IF(IN) INSIDE = 1
C      IF(.NOT.IN) INSIDE = 0
C      RETURN
C      END
C
C      SUBROUTINE COCRDS(XARR1,YARR1,XARR2,YARR2,IFLAG,
C      +      LENAR1,LENAR2)
C
C      SUBROUTINE ACCEPTS AN ARRAY OF COORDINATES DEFINING A SEGMENT
C      FROM A DIGITISER FILE AND IF NECESSARY REVERSES THEIR ORDER
C      TO ENSURE THAT A SEQUENCE OF SEGMENTS JOIN EACH OTHER
C      IN THE CORRECT ORDER
C
C      DIMENSION XARR1(100),YARR1(100),XARR2(500),YARR2(500)
C      J=LENAR2
C      IF (IFLAG.EQ.1) GOTO 100
C
C      COPY COORDINATES TO THE END OF 2ND ARRAYS
C
C      DO 9 I=1,LENAR1
C      J=J+1
C      XARR2(J) = XARR1(I)
C      YARR2(J) = YARR1(I)
C      9  CONTINUE
C      GOTO 199
C
C      TO REVERSE THE COORDINATES...
C
C      100 DO 8 I=LENAR1,1,-1
C      J=J+1
C      XARR2(J) = XARR1(I)
C      YARR2(J) = YARR1(I)
C      8  CONTINUE
C
C      199 LENAR2=J
C      RETURN
C      END

```

A2.2.5 (i) Program MPPTP4

(ii) Determines the boundaries of the smallest rectangles to enclose enumeration districts and allocates postcodes to these rectangles.

(iii) Input files.

a) Contains centroid coordinates of all postcodes.

b) Contains strings of coordinates defining ED boundaries.

Output files.

a) Contains a list of postcodes followed by all EDs within which each postcode might lie according to the smallest rectangle enclosing the ED. The position of the ED within the data file is also recorded.

(iv) Program description.

a) Main program.

The file of ED coordinates is read first and for each area the minimum and maximum values of x and y (i.e. the coordinates of the smallest enclosing rectangle) are held in a table. The centroid coordinates for each postcode are then read and compared with the values in the table. Each time the postcode is found to lie in a rectangle the ED name and its position in the data file are written to the output file MPRECT.

(v) Note.

a) This program is run in conjunction with program MPPTP5.

A2.2.6 (i) Program MPPTP5

(ii) Allocates postcodes to enumeration districts using a point-in-polygon routine.

(iii) Input files.

a) MPRECT-- output from MPPTP4.

Program MPPTP4

```

      PROGRAM MPPTP4(INPUT,OUTPUT,TAPE1,TAPE2,TAPE3)
C PROGRAM READS POSTCODE AND E.O. FILES AND CREATES FILE
C CONTAINING FOR EACH POSTCODE A LIST OF ALL E.O.S THE
C CODE COULD POSSIBLY LIE IN, BASED ON THE SMALLEST
C RECTANGLE ENCLOSING THE E.O.
C
      CHARACTER EDCODE*8
      DIMENSION XARR2(2500),YARR2(2500)
C
C SET UP ARRAY WHICH HOLDS XMAX,XMIN,YMAX,YMIN FOR
C GTR.LEICESTER
C
100   READ (1)EDCODE,LENAR2,(XARR2(K),YARR2(K),K=1,LENAR2)
      XMAX=0.0
      YMAX=0.0
      XMIN=999.99
      YMIN=999.99
      DO 9 K=1,LENAR2
      X=XARR2(K)
      Y=YARR2(K)
      IF (X.GT.XMAX)XMAX=X
      IF (X.LT.XMIN)XMIN=X
      IF (Y.GT.YMAX)YMAX=Y
      IF (Y.LT.YMIN)YMIN=Y
9     CONTINUE
      PRINT *, 'XMIN=',XMIN, 'XMAX=',XMAX
      PRINT *, 'YMIN=',YMIN, 'YMAX=',YMAX
C
C ESTABLISH WHETHER THE POSTCODE LIES WITHIN THE RECTANGLE
C ENCLOSING GTR LEICS, IF SO, WRITE IT TO FILE
C
102   READ(2,1000,ERR=103,END=9999)NUM,XPC,YPC
1000  FORMAT(I5,22X,2F4.1)
C
C ELIMINATE DUMMY POSTCODES
C
      IF (XPC.LT.425.OR.YPC.LT.275) GOTO 103
      IF(XPC.LE.XMAX.AND.XPC.GE.XMIN
+      .AND.YPC.LE.YMAX.AND.YPC.GE.YMIN)THEN
      WRITE(3,1004)NUM,XPC,YPC
1004  FORMAT (I5,2F5.1)
      ENDIF
      GOTO 102
C
C IF SPURIOUS POSTCODE READ, PRINT CODE NUMBER & CONTINUE
C
103   PRINT *,NUM
      GOTO 102
C
9999  STOP
      END

```


Program MPPTP5

```

PROGRAM MPPTP5 (INPUT,OUTPUT,TAPE1,TAPE2,TAPE3,TAPE4,
+ TAPE5=OUTPUT)
C
C PROGRAM TO TEST WHICH ENUMERATION DISTRICTS A SET OF
C POSTCODES LIE IN
C
C     DIMENSION XARR2(2500),YARR2(2500)
C
C READ THE COORDINATES FOR THE GTR LEICR BOUNDARY AND EACH PCCDE
C
100   READ (2) EDCCDE,LENAR2,(XARR2(K),YARR2(K),K=1,LENAR2)
106   READ (1,1000,END=9999) NUM,XPC,YPC
1000  FORMAT (I5,2F5.1)
      INSIDE=9
C
C CALL THE POINT-IN-POLYGON ROUTINE
C
C     CALL SROUND(XARR2,YARR2,LENAR2,XPC,YPC,INSIDE)
C
C TEST FOR WHERE THE POSTCODE LIES IN RELATION TO THE S.D.
C
      IF (INSIDE.EQ.0) WRITE (7,1001)NUM
      IF (INSIDE.EQ.1) WRITE (3,1001)NUM
      IF (INSIDE.EQ.2.OR.INSIDE.EQ.3) WRITE (4,1001)NUM
1001  FORMAT (I5)
      GOTO 106
9999  STOP
      END
C
C     SUBROUTINE SROUND(XARRAY,YARRAY,NPOINT,XTEST,YTEST,INSIDE)
C
C A SUBROUTINE TO TEST WHETHER (XTEST,YTEST) IS INSIDE THE POLYGON
C DEFINED BY THE NPOINT VERTICES (XARRAY(I),YARRAY(I)).
C
C INSIDE RETURNS THE VALUE 0 IF THE POINT IS OUTSIDE THE POLYGON
C                               1 IF IT IS INSIDE
C                               2 IF IT IS ON A SIDE OF THE POLYGON
C                               3 IF IT IS A VERTEX OF THE POLYGON
C
C THE TEST IS WHETHER AN ODD OR EVEN NUMBER OF SIDES OF THE POLYGON
C INTERSECT WITH A TEST LINE EXTENDING FROM THE TEST POINT VERTICALLY
C UPWARDS. THE LOGICAL VARIABLE IN IS INITIALLY FALSE AND IS
C NEGATED WHENEVER AN INTERSECTION IS FOUND, SO IF IT ENDS UP TRUE
C THE THE TEST POINT IS INSIDE THE POLYGON.
C
C IF THE TEST POINT IS FOUND TO BE ON A SIDE OR VERTEX, THERE IS
C NO NEED TO CONTINUE EXAMINING FURTHER SIDES OF THE POLYGON.
C
C A SIDE MEETS THE TEST LINE IF IT X-COORDINATES ARE ON EITHER SIDE
C OF XTEST AND ITS Y-COORDINATE AT THE POSITION XTEST IS GREATER THAN
C YTEST.
C
C AWKWARD CASES ARISE IF XTEST IS EQUAL AN X-COORDINATE OF A VERTEX.
C THIS CASE IS CONSIDERED ONLY WHEN THE X-COORDINATE IS THE SECOND
C ONE FOR THE SIDE: THIS ENSURES THAT THE CASE IS DEALT WITH ONLY ONCE
C FOR THAT SIDE.
C
C IF THE Y-COORDINATE IS LESS THAN YTEST THEN THERE IS NO INTERSECTION.
C IF THE Y-COORDINATE IS EQUAL TO YTEST THEN THE TEST POINT COINCIDES
C WITH A VERTEX.
C IF THE Y-COORDINATE IS GREATER THAN YTEST, THEN THE TEST LINE PASSES
C THROUGH A VERTEX.

```

```

C
C IN THIS CASE THERE IS AN INTERSECTION
C IF THE NEXT SIDE TO BE CONSIDERED LIES ON THE OPPOSITE SIDE OF THE TEST
C LINE FROM THE CURRENT SIDE. THAT WILL BE SO IF XTEST IS BETWEEN THE
C FIRST X-COORDINATE OF THE CURRENT SIDE AND THE NEXT X-COORDINATE TO BE
C CONSIDERED.
C IF THE NEXT X-COORDINATE LIES ON THE SAME SIDE OF THE TEST LINE AS THE
C CURRENT FIRST X-COORDINATE, THEN THERE IS NO INTERSECTION.
C IF THE NEXT X-COORDINATE IS EQUAL TO XTEST, THEN THE TEST POINT LIES
C ON THE NEXT SIDE, POSSIBLY PROJECTED, I.E. THE TEST LINE IS COLINEAR
C WITH A SIDE.
C
C IF THE Y-COORDINATE IS LESS THAN YTEST, THEN THE TEST POINT LIES ON A
C SIDE OF THE POLYGON.
C IF THE Y-COORDINATE IS EQUAL TO YTEST, THEN THE TEST POINT COINCIDES WITH
C A VERTEX OF THE POLYGON.
C IF THE Y-COORDINATE IS GREATER THAN YTEST, IT IS NECESSARY TO LOOK AHEAD
C UNTIL AN X-COORDINATE IS FOUND WHICH IS NOT COLINEAR WITH THE TEST LINE.
C THEN IF THAT X-COORDINATE IS ON THE OPPOSITE SIDE OF THE TEST LINE TO
C THE CURRENT FIRST X-COORDINATE, THERE IS AN INTERSECTION. IF IT IS ON
C THE SAME SIDE, THERE IS NO INTERSECTION. AFTER LOOKING AHEAD FOR SO
C MANY SIDES, THE NORMAL TESTING IS RESUMED WITH THE NEXT SIDE AFTER THAT
C JUST LOOKED AT.
C
C SINCE THE FIRST SIDE OF THE POLYGON TO HAVE BEEN TESTED MAY HAVE BEEN
C WITHIN A SEQUENCE OF SIDES COLINEAR WITH THE TEST LINE, IN LOOKING
C AHEAD IT MAY BE NECESSARY TO CONTINUE NOT JUST TO THE LAST SIDE BUT
C TO LOOK AGAIN AT ALL THE SIDES FROM THE FIRST UP TO THE CURRENT ONE.
C IF LOOKING AHEAD DOES NOT RESOLVE THE QUESTION OF WHETHER THERE IS AN
C INTERSECTION, THEN WE HAVE THE CASE WHERE ALL SIDES ARE COLINEAR WITH THE
C TEST LINE AND DO NOT MEET THE TEST POINT. I.E. THE "POLYGON" IS A
C VERTICAL LINE, AND HENCE THE TEST POINT IS NOT INSIDE IT.
C
      REAL XARRAY(NPCINT), YARRAY(NPOINT)
      LOGICAL IN
C
C UNTIL AN INTERSECTION IS FOUND ASSUME THE TEST POINT IS NOT INSIDE
C THE POLYGON.
C
      IN = .FALSE.
C
C GO THROUGH ALL VERTICES OF THE POLYGON
C
      NP = 1
20    X1 = XARRAY(NP)
      Y1 = YARRAY(NP)
      X2 = XARRAY(MOD(NP,NPOINT)+1)
      Y2 = YARRAY(MOD(NP,NPOINT)+1)
C
C TEST WHETHER XTEST LIES BETWEEN X1 AND X2
C
      BETWEN = (XTEST-X1)*(XTEST-X2)
C
C IF IT DOES NOT, THEN GO ON TO THE NEXT SIDE OF THE POLYGON
C
      IF(BETWEN.GT.0) GOTO 10
      IF(BETWEN.EQ.0) GOTO 40
C
C IN THIS CASE VERTICES ARE ON EITHER SIDE OF THE TEST LINE.
C SEE WHETHER THE POINT OF INTERSECTION OF THE SIDE WITH THE
C TEST LINE IS ABOVE OR BELOW THE TEST POINT.
C

```

```

      YINT = Y1 + (XTEST-X1)*(Y2-Y1)/(X2-X1)
C
C IF THE INTERSECTION IS BELOW THE TEST POINT, THEN GO ON
C TO THE NEXT SIDE OF THE POLYGON
C
      IF(YINT.LT.YTEST) GOTO 10
      IF(YINT.EQ.YTEST) GOTO 50
C
C THE SIDE MEETS THE TEST LINE ABOVE THE TEST POINT:
C COUNT THIS INTERSECTION.
C
      IN = .NOT.IN
      GOTO 10
C
C THE MEETING POINT COINCIDES WITH THE TEST POINT, SO THE
C TEST POINT IS ON A SIDE OF THE POLYGON
C
      50  INSIDE = 2
          RETURN
C
C THE TEST LINE PASSES THROUGH A VERTEX OF THE POLYGON.
C IF THE VERTEX IS THE FIRST ONE OF THE CURRENT SIDE,
C THE PROBLEM HAS ALREADY BEEN DEALT WITH (OR WILL BE
C EVENTUALLY IF IT IS THE FIRST VERTEX OF THE POLYGON)
C AND SO WE GO ON TO CONSIDER THE NEXT VERTEX.
C
      40  IF(XTEST.NE.X2) GOTO 10
C
C IF THE VERTEX IS BELOW THE TEST POINT, THERE IS NO
C INTERSECTION AND WE GO ON.
C
      IF(Y2.LT.YTEST) GOTO 10
      IF(Y2.NE.YTEST) GOTO 70
C
C THE TEST POINT COINCIDES WITH A VERTEX
C
      INSIDE = 3
      RETURN
C
C THE TEST LINE PASSES THROUGH A VERTEX. WE NOW LOOK
C AHEAD TO SUBSEQUENT SIDES TO SEE IF THIS IS TO COUNT
C AS AN INTERSECTION.
C
      70  DO 110 NPX = NP+2, NP+NPOINT
          X3 = XARRAY(MOD(NPX-1, NPOINT)+1)
          BETWEN = (XTEST-X1)*(XTEST-X3)
C
C IF X1 AND X3 LIE ON THE SAME SIDE OF XTEST, THERE
C IS NO INTERSECTION
C
      IF(BETWEN.GT.0) GOTO 10
      IF(BETWEN.EQ.0) GOTO 80
C
C XTEST LIES BETWEEN X1 AND X3 SO THERE IS AN INTERSECTION
C
      IN = .NOT.IN
C
C THE SIDES LOOKED AT NEED NOT BE EXAMINED AGAIN
C
      NP = NPX
      GOTO 10
C

```



```

C IN THIS CASE THE TEST POINT IS COLINEAR WITH A SIDE OF
C THE POLYGON
C
80  Y3 = YARRAY(MOD(NPX-1,NPOINT)+1)
    IF(Y3.GE.YTEST) GOTO 90
C
C THE SIDE EXTENDS BELOW THE TEST POINT, AND HENCE
C PASSES VERTICALLY THROUGH IT.
C
    INSIDE = 2
    RETURN
90  IF(Y3.NE.YTEST) GOTO 110
C
C THE TEST POINT COINCIDES WITH A VERTEX
C
    INSIDE = 3
    RETURN
C
C THE TEST LINE IS COLINEAR WITH A SIDE. IT IS NOT YET
C DETERMINED WHETHER THIS COUNTS AS AN INTERSECTION, SO
C CONTINUE LOOKING AHEAD
C
110  CONTINUE
C
C IF WE HAVE LOOKED AHEAD THROUGH ALL VERTICES AGAIN AND
C NOT SETTLED WHETHER THERE IS AN INTERSECTION, THEN ALL
C SIDES MUST BE COLINEAR WITH THE TEST LINE BUT NONE PASS
C THROUGH THE TEST POINT. THEREFORE THE TEST POINT IS
C OUTSIDE THE LINEAR "POLYGON".
C
    INSIDE = 0
    RETURN
C
C END OF LOOP. GO TO NEXT VERTEX, IF ANY.
C
10  NP = NP+1
    IF(NP.LE.NPOINT) GOTO 20
C
C THE TEST POINT DOES NOT LIE ON A SIDE OR VERTEX OF THE
C POLYGON. IN HAS RECORDED WHETHER THE TEST POINT IS INSIDE
C OR OUTSIDE THE POLYGON.
C
    IF(IN) INSIDE = 1
    IF(.NOT.IN) INSIDE = 0
    RETURN
    END
C
    SUBROUTINE COCRDS(XARR1,YARR1,XARR2,YARR2,IFLAG,
+      LENAR1,LENAR2)
C
C SUBROUTINE ACCEPTS AN ARRAY OF COORDINATES DEFINING A SEGMENT
C FROM A DIGITISER FILE AND IF NECESSARY REVERSES THEIR ORDER
C TO ENSURE THAT A SEQUENCE OF SEGMENTS JOIN EACH OTHER
C IN THE CORRECT ORDER
C
    DIMENSION XARR1(100),YARR1(100),XARR2(500),YARR2(500)
    J=LENAR2
    IF (IFLAG.EQ.1) GOTO 100
C
C COPY COORDINATES TO THE END OF 2ND ARRAYS
C
    DO 9 I=1,LENAR1

```

```

      J=J+1
      XARR2(J) = XARR1(I)
      YARR2(J) = YARR1(I)
9     CONTINUE
      GOTO 199

C
C TO REVERSE THE COORDINATES...
C
100   DO 8 I=LENAR1,1,-1
      J=J+1
      XARR2(J) = XARR1(I)
      YARR2(J) = YARR1(I)
      8   CONTINUE
C
199   LENAR2=J
      RETURN
      END

```

b) Contains strings of coordinates defining ED boundaries.

Output files.

a), b) and c) as program MPPTPL.

(iv) Program description.

a) Main program.

Reads postcodes and ED names from MPRECT and uses the pointers to the second input file to read only those records within which the postcode might lie. As each ED boundary is read subroutine SROUND is called to establish the relationship of postcode to ED.

(v) Note.

a) This program is run in conjunction with program MPPTP4.

A2.3 Programs for data display.

A2.3.1 (i) Program MPPLLOT

(ii) Plots the Greater Leicester boundary and plots symbols at enumeration district centroids proportional to data values.

(iii) Input files.

a) Contains coordinates of Greater Leicester boundary.

b) Contains selected census data.

Output is sent to the plotter.

(iv) Program description.

a) Main program.

Dimensions and initialises variables. Reads boundary coordinates and census data. Calculates minimum, maximum, mean and missing values. Converts data to appropriate values for plotting. Controls plotting.

The following subroutines from the GHOST80 graphics package are accessed:

PAPER Initialises GHOST80

GPSTOP Sets limit to graphical output

FRAME Select new sheet of plotting paper

Program MPPLLOT

```

/JO3
MPJPLOT.
/USER
CHARGE,GEOG,GGPH.
FTNS.
GET,TAPE1=MPUNFBD/PN=SCRATCH,NA.
GET,TAPE2=MPDPMS/PN=SCRATCH,NA.
ATTACH,GHOST80/UN=LIBS.
LIBRARY,GHOST80.
LGO.
ATTACH,GRIDCAL/UN=LIBS.
GRIDCAL.
/EOB

      PROGRAM MPPLOTO
C
C PROGRAM PLOTS GREATER LEICESTER BOUNDARY AND SYMBOLS AT
C E.C. CENTROIDS PROPORTIONAL TO VARIABLE VALUES
C
C COORDS OF BOUNDARY
C
      DIMENSION X1(2000),Y1(2000)
C
C CENTROIDS OF P.M.S. OBSERVATIONS
C
      DIMENSION XCENT(1400),YCENT(1400),N2(1400)
      DIMENSION VALUE(1400),NUM(1400)
C
C MATRIX OF ALL VARIABLES FOR ALL P.M.S. OBSERVATIONS
C
      CHARACTER FORM*20,AREA*8
C
C INITIALISE
C
      NUMOBS=1369
      VMISS=99.0
C
C READ IN BOUNDARY COORDINATES
C
      READ(1)AREA,LENAR,(X1(I),Y1(I),I=1,LENAR)
C
C READ IN COORDINATES & PMS VARIABLE
C
      DO 9 J=1,NUMOBS
      READ (2,2000) NUM(J),XCENT(J),YCENT(J),VALUE(J)
2000  FORMAT(I3,17X,2F4.1//73X,=1.0/////))
9      CONTINUE
C
C COMMENCE PLOTTING
C INITIALISE GHOSTS0
C
      CALL PAPER(1)
C
C DEFINE LIMITS OF PLOTTING SURFACE
C
      CALL PSPACE(0.1,0.633,0.1,0.751)
      CALL MAP(450.0,468.0,293.0,315.0)
C
C PLOT OUTLINE
C
      DO 8 K=1,LENAR
      X=X1(K)
      Y=Y1(K)

```

```

      IF (K.EQ.1) THEN
        CALL PCINT(X,Y)
      ELSE
        CALL JGIN(X,Y)
      ENDIF
8     CONTINUE
C
C PLOT BORDER
C
      CALL BORDER
C
C
C PLOT SYMBOL AT EACH CENTROID
C
      DO 5 J=1,NUMCBS
        IF(NUM(J).LT.500) THEN
          ISYMSL=227
        ELSE
          ISYMSL=229
        ENDIF
        IF(VALUE(J).EQ.VMISS)GOTO 5
        IF(VALUE(J).NE.1.0) GOTO 5
        SIZE=0.45
        X=XCENT(J)
        Y=YCENT(J)
        CALL CTRSIZE(SIZE)
        CALL PLOTNC(X,Y,ISYMSL)
5     CONTINUE
C
C PLOT KEY SYMBOLS
C
      CALL CTRSIZE(0.45)
      CALL PLOTNC(467.25,294.55,227)
      CALL PLOTNC(467.25,293.75,229)
C
C TERMINATE
C
      CALL GREND
      STOP
      END

```

MAP Define mathematical space
 POINT Moves pen and marks paper with a dot
 JOIN Draws straight line to new point
 POSITN Moves current plotting position
 CTRS12 Sets character size in mathematical space
 THICK Draws multiple lines to increase character
 thickness
 TYPECS Plots characters
 REDPEN Select red plotting pen
 BLKPEN Select black plotting pen
 PLOTNC Plots single character
 TYPENF Plots real number
 TYPENI Plots integer
 GREND Terminates plotting.

A2.3.2 (i) Program MPBOX

(ii) Plots box plots and dispersion diagrams for perinatal mortality data.

(iii) Input file.

a) Contains perinatal data.

Output is sent to the plotter.

(iv) Program description.

a) Main program.

Dimensions and initialises variables. Reads in data and eliminates missing values. Sorts data values into ascending sequence, calculates minimum, maximum, median, quartiles and range. Controls plotting.

The following subroutines from the GHOST80 graphics package are accessed:

PAPER Initialises GHOST80

FRAME Select new sheet of plotting paper

Program MPBOX

```

PROGRAM MPBOX
C
C PROGRAM PLCTS BOX PLOT AND DISPERSION DIAGRAM FOR CENSUS DATA
C
      CHARACTER*22 TITLE(5)
      DIMENSION VALUE(227),VAR(227,5)
      NUMAR=227
C
C READ IN DATA
C
      DO 9 I=1,NUMAR
        READ(1,1000)(VAR(I,J),J=1,5)
1000    FORMAT(8X,5F10.4)
        CONTINUE
9
C
C SET UP TITLES
C
      TITLE(1)='CHILDREN AGED 0-4'
      TITLE(2)='PER TOTAL RESIDENTS'
      TITLE(3)='PER TOTAL AREA'
      TITLE(4)='PER RESIDENTS PER AREA'
      TITLE(5)='CHI-SQUARE VALUES'
C
C PLCT BOX PLOT & DISPERSION DIAGRAM FOR EACH VARIABLE
C
      DO 7 K=1,5
C
C READ VALUES FROM TABLE INTO ARRAY
C
        DO 6 L=1,NUMAR
          VALUE(L)=VAR(L,K)
6        CONTINUE
C
C PLOT USING GHOST80
C
C INITIALISE
C
        CALL PAPER(1)
C
C SORT THE VALUES INTO ASCENDING SEQUENCE USING N.A.G. ROUTINE
C
        LOWER=1
        IUPPER=NUMAR
        IFAIL=0
        CALL M01ANF(VALUE,LOWER,IUPPER,IFAIL)
C
C CALCULATE MIN,MAX,MEDIAN,QUARTILES & RANGE
C
        VALMIN=VALUE(LOWER)
        VALMAX=VALUE(IUPPER)
        HALF=IUPPER/2
        QUART=IUPPER/4
        MIDDLE=IFIX(HALF)
        VALMED=VALUE(MIDDLE)
        IUPQUR=IFIX(IUPPER-(QUART))
        LOQUAR=IFIX(QUART)
        VALUPQ=VALUE(IUPQUR)
        VALLOC=VALUE(LOQUAR)
        RANGE=VALUE(IUPPER)-VALUE(LOWER)
        TOPAXS=VALMAX+(RANGE/20.0)
        BOTAXS=VALMIN-(RANGE/20.0)
        CALL FRAME

```

```

      CALL PSPACE(0.1,0.85,0.1,0.35)
      CALL MAP(0.0,1.0,BCTAXS,TOPAXS)
C
C DRAW SCALE
C
      CALL YAXIS
C
C DRAW BOX PLOT
C
      CALL PLOTNC(0.1,VALMED,161)
      CALL PLOTNC(0.1,VALMAX,162)
      CALL PLOTNC(0.1,VALMIN,162)
      CALL BRCKEN(10,5,10,5)
      CALL POSITN(0.1,VALMAX)
      CALL JOIN(0.1,VALUPQ)
      CALL POSITN(0.1,VALLOQ)
      CALL JOIN(0.1,VALMIN)
      CALL BRCKEN(0,0,0,0)
      CALL POSITN(0.075,VALUPQ)
      CALL JOIN(0.125,VALUPQ)
      CALL JOIN(0.125,VALLOQ)
      CALL JOIN(0.075,VALLOQ)
      CALL JOIN(0.075,VALUPQ)
C
C DRAW DISPERSION DIAGRAM
C
      DO 8 I=1,NUMAR
      IF(I.EQ.1)THEN
        CALL POSITN(0.2,VALUE(I))
        CALL LINE(0.01,0.0)
      ELSE
        IF(VALUE(I).EQ.VALUE(I-1))THEN
          CALL MOVE(0.005,0.0)
          CALL LINE(0.01,0.0)
        ELSE
          CALL POSITN(0.2,VALUE(I))
          CALL LINE(0.01,0.0)
        ENDIF
      ENDIF
      CONTINUE
8
C
C DRAW TITLE AND ANNOTATION
C
      CALL CSPACE(0.1,0.95,0.85,0.91)
      CALL PLACE(1,1)
      CALL CTRMAG(3)
      CALL CRLNFD
      CALL TYPECS('MEDIAN = ',9)
      CALL TYPENF(VALMED,3)
      CALL CRLNFD
      CALL TYPECS('MAXIMUM VALUE = ',16)
      CALL TYPENF(VALMAX,3)
      CALL CRLNFD
      CALL TYPECS('MINIMUM VALUE = ',16)
      CALL TYPENF(VALMIN,3)
      CALL CSPACE(0.0,0.0,0.0,0.0)
      CALL CTRMAG(10)
C
C
7      CONTINUE
C
C TERMINATE
C

```

CALL GRENO
STOP
END

PSPACE	Define physical plotting space
MAP	Define mathematical space
YAXIS	Draws y axis
PLOTNC	Plots character
BROKEN	Draws broken line
POSITN	Moves current plotting position (absolute movement)
JOIN	Draws straight line to new point (absolute coordinates)
LINE	Draws straight line to new point (relative coordinates)
MOVE	Moves current plotting position (relative movement)
CSPACE	Maps character space within physical space
PLACE	Set current position for plotting characters
CTRMAG	Sets character magnification
CRLNFD	Sets current character position to start of next line
TYPECS	Plots characters
TYPENF	Plots real number
TYPENI	Plots integer
GREND	Terminates plotting

The following subroutine from the NAG library is accessed:

M01ANF Sorts data values into ascending sequence.

NMAPS 1NAREA238
NDRAW238SHOAL 4INTVL OPRINT OTEXT 1
ARSEL 0ADLIN UKEY8X OMSIZE 3
WINDO 449.0 469.0 292.0 315.0
SIZE 230.0
SHADCOLOUR 1
VALUE

7A
7B
7C
7C
7E
7F
7G
7H
7I
7J
7K
7L
7M
7N
7O
7P
7Q
7R
7S
7T
7U
7V
7W
7X
7Y
7Z
7AA
7BB
7CC
7CC
7EE
7FF
7GG
7HH
7II
7JJ
7KK
7LL
7MM
7NN
7CC
7PP
4A
4B
4C
4C
4E
4F
4G
4H
4I
4J
4K
4L
4M
4N

4C
4D
4E
4R
4S
4T
4U
4V
4W
4X
4Y
1A
1B
1C
1D
1E
1F
1G
1H
1I
1J
1K
1L
17A
17B
17C
17D
17E
17F
17G
14A
14B
14C
14E
14F
14G
14H
14I
14J
14K
14L
14M
14N
14C
14P
14Q
14R
14S
14T
14U
14V
14W
14X
14BB
14CC
14DD
14EE
14FF
14GG
14HH
14JJ
14KK

14LL
14MM
14NN
11A
11B
11C
11D
11E
11F
11G
11H
11I
11J
11K
11L
11M
11N
11O
11P
11Q
11R
11S
11T
11U
11V
11W
11X
11Y
27A
27B
27C
27C
27E
27F
27G
27H
27I
27J
27K
27L
27M
27N1
27N2
27N3
27P
27Q
27R
27S
27V
27W
27X
27Y
27Z
27AA
27BB
27CC
24A
24B
24C
24D
24E
24F

24G
 24H
 24I
 24J
 24K
 24L
 24M
 24N
 24O
 24P
 24Q
 24R
 24S
 24T
 24U
 24V
 24W
 24X
 24Y
 24Z
 24AA
 24BB
 24CC
 24CD
 24EE
 24FF
 24GG
 24HH
 24II
 21A
 21B
 21C
 21D
 21E
 21F
 21G
 21H
 21I
 21J
 21K
 21L
 21M
 21N
 21O
 21P
 21Q
 21S
 21T
 21U
 7
 4
 1
 17
 14
 11
 27
 24
 21

PLCEO	449.5	314.0	2	2
9 CLUSTERS - CLUSTER 7N SPLIT#.				
PLCEO	455.0	297.7	0	1
CLUSTER 1#.				

PLCE0	465.0	297.1	0	1
CLUSTER 2#.				
PLCE0	465.0	296.5	0	1
CLUSTER 3#.				
PLCE0	465.0	295.9	0	1
CLUSTER 4#.				
PLCE0	465.0	295.3	0	1
CLUSTER 5#.				
PLCE0	465.0	294.7	0	1
CLUSTER 6#.				
PLCE0	465.0	294.1	0	1
CLUSTER 7#.				
PLCE0	465.0	293.5	0	1
CLUSTER 8#.				
PLCE1	465.0	292.9	0	1
CLUSTER 9#.				

Appendix 3. The Perinatal Mortality Survey

A3.1 The Questionnaire

CONFIDENTIAL

PERINATAL SURVEY

This questionnaire is to be completed for stillbirths and deaths within seven days of life, the day of birth being taken as the first day. In cases of multiple pregnancy when more than one infant is affected, a separate questionnaire is to be completed for each.

Mother's name: Surname_____

Forenames_____

Address_____

[N.B. This information was not made available to the current project]

For office use only:

Survey number_____

Year of Survey_____

Unit number of mother's case notes_____

How to complete the questionnaire

General

1. A questionnaire is to be completed for all stillbirths and deaths within seven days of birth. Separate questionnaires to be completed for each infant in a multiple pregnancy.
2. The questionnaire has two parts; both are to be completed before the mother leaves the unit, or ceases to be visited by the person responsible for her care during the birth of the infant.
3. Part I to be completed first using all the case notes and the cooperation card.
4. Part II is to be completed by interviewing the patient within a week of delivery. A simple explanation should be offered about the survey and the patients' cooperation sought. Assurances are to be given about the confidentiality and eventual anonymity of the data that is collected.
5. On completion the questionnaire is to be returned to the Department of Community Health, University of Leicester.

Definitions to be used for questions 39 and 40.

Occupation

The occupation of a person is the kind of work he or she performs, the nature of the factory, business or service in which the person is employed having no bearing on his occupation. Example: a crane driver may be employed in a shipyard, engineering works or in building construction but this has no bearing on his occupation. All crane drivers being from the same occupational group.

Industry

The industry in which an individual is engaged is determined, whatever his occupation, by reference to the business or economic activity in which his occupation is followed. Example: the man who is occupationally a carpenter or car man is classified industrially to building if employed by a builder, but to brewing if employed by a brewer.

Part I - Obstetric Notes

1. Ward UD/RD/MB _____
Place of delivery _____
Intended place of delivery _____
Survey No. of matching case or control _____

Indicate type of case or control:

Control

Singleton birth

Multiple birth

Case

Singleton birth

Single case in multiple birth

Multiple case (first)

Multiple case (subsequent)

If multiple birth specify:

Multiplicity of birth _____

No. of perinatal deaths _____

Mother's personal details:

2. Date of birth Day _____ Month _____ Year _____
Age _____

3. Marital state Married
 Single
 Widowed
 Separated
 Divorced

If single, widowed, separated or divorced, was the mother cohabiting during the pregnancy? YES/NO

Mother's general medical history:

4. Give details of past medical history including blood transfusions which had occurred before the pregnancy being investigated.
List specific diseases: _____
Nil Apparent

5. Give details of the present medical history since the beginning of the pregnancy under review (including any blood transfusions)

Nil Apparent

Obstetric history:

6. Using the convention (live births + stillbirths) + (abortions) write the number of such births the woman has had, including this pregnancy.
(e.g. 4 + 2) [N.B. Count births not pregnancies].

Interval between previous and present pregnancy _____

7. Give details of the obstetric history EXCLUDING the pregnancy under review, but INCLUDING previous stillbirths and abortions. Under infant outcome include any malformations. If the infant is dead state age at death. Indicate multiple pregnancies and treat each infant as a separate pregnancy.

Pregnancy 1: Year_____

Gestation_____

Antenatal complications_____

Type of delivery_____

Complications of delivery_____

Complications of puerperium_____

Birth weight of infant_____

Sex_____

Infant outcome_____

(repeat for all previous pregnancies)

8. Care during pregnancy:

A. Planned care of the pregnancy:

- (i) Full consultant care
- (ii) Shared consultant/G.P. care
- (iii) Full G.P./midwife care
- (iv) None

Initial intended place of delivery_____

B. Change in care during the pregnancy:

- Transfer to (i) Full consultant care
- (ii) Shared consultant/G.P. care
- (iii) Full G.P./midwife care
- (iv) No change

C. Change in care during labour:

- Transfer to (i) Consultant unit
- (ii) Shared consultant/G.P. care
- (iii) Full G.P./midwife care
- (iv) Unattended delivery
- (v) No change

Antenatal Care:

9. Was antenatal care given? YES/NO

If yes, what was the duration of pregnancy in weeks at the first recorded contact? (i.e. by doctor or midwife)_____

10. How many attendances for outpatient antenatal care were recorded on both the hospital and general practitioner/midwifery records

Before 20 weeks gestation_____

20 weeks gestation and after _____

11. How many days during the antenatal period did the mother receive in-patient care? (include day of admission)_____

Pregnancy being reviewed:

12. Date of first day of last normal menstrual period -

Day_____ Month_____ Year_____

Estimated date of delivery - Day_____ Month_____

Year_____

Length and duration of menstrual cycle _____ days

Example 5/28 days

Is the patient certain of the date of her last menstrual period? YES/NO

13. State mother's height as feet and inches

or cms.

14. Give the result of the first recorded haemoglobin estimation as gm/100mls of blood _____
15. Give the result of the first recorded diastolic pressure_____
16. Were antibodies stated to be present in the first blood sample recorded for the pregnancy under review?
If 'Yes' specify:_____
17. Give the results, and the number of weeks gestation, of the first and second recorded alpha feto-protein levels in ug/litre of blood.
1st result_____at_____weeks
2nd result_____at_____weeks
18. Was amniocentesis performed? YES/NO
If so, when (in weeks gestation)_____weeks
Results x FP_____µg/litre
Chromosomes:_____

Length of pregnancy at labour and delivery:

19. How many weeks pregnant was the mother thought to be at the time of labour and delivery?_____

Presentation of foetus:

20. Indicate presentation of foetus:

Cephalic
Breech
Other

21. Did the cord prolapse? YES/NO

Type of labour:

22. Indicate the type of labour:

Spontaneous
Surgical induction
Medical induction - prostaglandins
Medical induction - oxytocics
Augmented labour - ARM
Augmented labour - oxytocics
Augmented labour - ARM and O
Other induction
Did not labour

Mode of delivery:

23. Indicate the type of delivery the mother experienced:

Normal
Forceps
Ventouse
Caesarian
Breech extraction

24. List any drugs given to the mother during induction, labour and delivery.

Name of drug_____

Dose_____

No. of times given_____

Give the time taken to complete the following:

- A. First stage of labour Hrs._____Mins._____
B. Second stage of labour Hrs._____Mins._____
C. Did not labour

Note: if the mother did not complete both stages of labour, please give time taken before intervention took place.

Epidural Anaesthetic:

Was an epidural anaesthetic given? YES/NO

If 'Yes' is the epidural form filed in the mother's record YES/NO

General Anaesthesia:

25. Indicate whether the mother had a general anaesthetic administered during part of her pregnancy, labour and puerperium:

No anaesthetic
Antenatal period
1st stage of labour
2nd stage of labour
3rd stage of labour
Post partum: (i) sterilization
(ii) other

Placenta and cord:

26. Indicate the weight of the placenta_____grammes

Were any abnormalities noted, including retroplacental clot? YES/NO

If 'Yes' what were they?_____

Were any abnormalities noted about the cord? YES/NO

If 'Yes' what were they?_____

Details of infant:

28. Total number of infants born during the pregnancy under review_____

29. Total number of infants from the pregnancy either stillborn or dead within seven days of birth_____

30. Infant's weight at birth_____grammes

31. Sex of infant_____

32. List Agpar score readings for live births:

A. 1 minute_____

B. 5 minutes_____

If stillbirth tick

33. Date of birth Date_____Month_____Year_____
Time of birth Hrs._____Mins._____(24 hour clock)

34. If the infant was a stillbirth, in which stage of pregnancy did the death occur?

Antenatal period
1st stage of labour
2nd stage of labour

35. For neonatal death indicate day of death. Days terminate at 12.00 midnight.
- Day 1*
 - Day 2
 - Day 3
 - Day 4
 - Day 5
 - Day 6
 - Day 7

*Each day represents an additional 24 hours of life.

Date of death Date_____ Month_____ Year_____

Time of death Hrs._____ Mins._____ (24 hour clock)

36. For a perinatal death, was a postmortem carried out? YES/NO

Part II - Maternal Interview

G.P. Details:

37. Mother's family doctor/clinic address:

Name_____

Address_____

Postcode_____

Antenatal clinic provided at surgery/practice? YES/NO

When did the mother say she first contacted the G.P. for antenatal care?
(Weeks gestation)_____

38. Marital state
- Married
 - Single
 - Widowed
 - Separated
 - Divorced

If single, widowed, separated or divorced, was the mother cohabiting during the pregnancy? YES/NO

Occupation

If the mother was unmarried and still receiving full-time education at school, turn to Question 40.
For the remainder, continue with Question 39.

39. Did you have an occupation other than home duties while pregnant? YES/NO
- If 'No' turn to Question 40.
- If 'Yes', what was your occupation?_____

In which industry did you work?_____

Were you self-employed? YES/NO

If 'Yes', how many employees did you have_____

If 'No', were you an employee? YES/NO

One of the following? Manager

 Forewoman/Supervisor

 Trainee

If a manager, how many employees did you care for?_____

During which month of your pregnancy did you finish work?

1	2	3	4	5	6	7
8	9					

Question 40 refers to the occupation of:

- A. The father of an unmarried school girl
- B. The husband of a married woman
- C. The partner of a cohabitee

If the mother did not have a regular partner and does not fit into category A, B or C, ignore question and continue with Question 41.

40. What was the occupation of your father/husband/partner while you were pregnant? _____

In which industry did he work?_____

Was he self-employed? YES/NO

If 'Yes', how many employees did he have? _____

If 'No', was he an employee? YES/NO

If an employee, was he one of the following?

Manager

Foreman/Supervisor

Trainee

If he was a manager, how many employees did he care for?_____

41. List any drugs you have been taking during the pregnancy until labour or induction, indicating the trimester(s) in which they were taken.

[illegible]

42. In which country were you born? _____

Using observation plus conversational 'probing', ascribe the mother to an ethnic group:

U.K.

Eire

Indian

Pakistani/Bangladeshi

Other Asian

African

West Indian

Other

(state _____)

43. What is your religion?_____

44. How many years have you been living in the United Kingdom?_____

45. How many years have you been living in Leicester(shire)?_____

46. Have you ever smoked as much as one cigarette a day for as long as a year?
YES/NO

Between the date of your last menstrual period and your delivery did you smoke as much as one cigarette a day? YES/NO

If 'Yes', during what period of the pregnancy?

Month: 1 2 3 4 5 6 7 8 9

On average how many cigarettes (or equivalent) per day?_____

(One small cigar = 2 cigarettes, one large cigar = 5 cigarettes, loz pipe tobacco = 28 cigarettes)

Alcohol:

Do you drink alcohol? (prompt: special occasion, home brewed beer or wine etc.) YES/NO

Did you drink at all during this pregnancy? (prompt: special occasions, before pregnancy was confirmed etc.) YES/NO

If so, during which period of the pregnancy?

Month:	1	2	3	4	5	6	7	8	9
During the first 3 months of your pregnancy, how often did you consume the following alcoholic drinks?									
	5 or more days per week	3-4 days per week	1-2 days per week	1-2 days per month	Less than once a month	Never	How much of this type of drink would you consume on a typical occasion?		
Shandy									
Beer, lager, stout or cider								pints	
Spirits or liqueurs, (e.g. rum, vodka, gin, whisky, advocaat etc)								pints	
Sherry or Martini (including Vermouth, port, Cinzano etc)								singles	
Wine (including champagne and Babycham)								glasses	

(Tick appropriate box)

During the 3 months before you came into hospital how often did you consume the following alcoholic drinks?	5 or more days per week	3-4 days per week	1-2 days per week	1-2 days per month	Less than once a month	Never	How much of this type of drink would you consume on a typical occasion?
Shandy							
Beer, lager, stout or cider							pints
Spirits or liqueurs (e.g. rum, vodka, gin, whisky, advocaat etc)							pints
Sherry or Martini (including Vermouth, port, Cinzano etc)							singles
Wine (including champagne and Babycham)							glasses

Have you been to any parties, weddings etc. where you've drunk more than usual? YES/NO
If so, please give details of the amount drunk and the approximate month of the pregnancy_____

(give details of ALL occasions)_____

How many cups of coffee did you drink (on average) per day during your pregnancy?_____

Diet:

How often did you eat the following during your pregnancy?

	5 or more times/week	3-4 times /week	1-2 times /week	Less than once/week	Never
Meat, (beef, mutton, lamb, pork)					
Poultry (chicken, duck etc.)					
Offal (liver, kidneys, heart etc)					
Other (sausages, beefburgers, faggots etc.)					
White fish (cod, coley, haddock etc.)					
Oily fish (sardines, mackerel, pilchards, herrings)					
Eggs					
Cheese					
Fresh fruit					
Yoghurt					

How often did you eat the following during your pregnancy?

	Regularly*	Occasionally	Never
Beans (includes: dhal, lentils, peas, baked beans etc.)			
Asian vegetables (includes: okra, bindi, karella etc.)			
Green vegetables (cabbage etc.)			
Root vegetables (carrots, turnips etc.)			
Potatoes			
Bread			
Breakfast cereals			
Chappatis			
Paratha			
Pouris			
Rice			
Pasta			
Crisps			
Nuts			
Cakes/biscuits			
Ghatia/chevda			
Sweets/chocolates			
Butter			
Margarine			
Oil			
Ghee			

*Regularly means at least 3-4 times per week.

How much milk did you drink per day during this pregnancy? _____ pints

Was the milk pasteurised or sterilised? _____

Did you boil ALL the milk? YES/NO

Did you have any vomiting during this pregnancy? YES/NO
If 'Yes', which months?

1 2 3 4 5 6 7 8 9

Did you have any diarrhoea during this pregnancy? YES/NO
If 'Yes', which months?

1 2 3 4 5 6 7 8 9

Did you have any loss of appetite during this pregnancy? YES/NO
If 'Yes', which months?

1 2 3 4 5 6 7 8 9

FOR OFFICE USE ONLY

47. List causes of infant's death:

I Direct cause

State foetal or maternal condition directly causing death.

(a) _____

Antecedant causes

State foetal and/or maternal conditions if any, giving rise to the above cause, stating the underlying cause last

due to (b) _____

due to (c) _____

II Other significant conditions of foetus or mother which may have contributed to but, in so far as is known, were not related to direct cause of death.

A3.2 Modifications to the Perinatal Mortality Survey Questionnaire

New questions on:

- | | |
|------|--|
| 1977 | Time of birth
Date of death
Time of death
Smoking habits
Control/case matching |
| 1978 | Area code
Intended place of delivery
Multiplicity (type of case or control)
Number of perinatal deaths
Cord prolapse |
| 1981 | Day of admission included for in-patient antenatal
care count
Alpha and feto-protein levels
Amniocentesis performed
Infant sex
Place of death
Details of the mother's G.P.
Mother's marital status
Month of pregnancy the mother finished work
Religion
Food and alcohol consumption
Vomiting, diarrhoea and loss of appetite |

Also a large number of minor modifications to existing questions, such as the addition of new options for certain answers.

Appendix 4. Principal Components Analysis : Results

A4.1 Eigenvalues : First 10 Components

Component 1	7.689
Component 2	6.031
Component 3	4.135
Component 4	2.735
Component 5	1.752
Component 6	1.686
Component 7	1.134
Component 8	0.884
Component 9	0.818
Component 10	0.761

A4.2 Component Loadings

Variable	Component 1	Component 2	Component 3	Component 4
DYNGCH	.27580	.62483	-.11597	.18624
DOLDCH	-.14214	.66420	-.44125	-.10160
HHSING	.55053	-.54499	.46775	.19086
HHCHIL	.47181	.61283	-.09550	-.26804
HHLPAR	.50832	-.20174	-.17162	.37190
HHXFAM	.11507	.72066	-.09648	-.49412
HSOWN	-.72561	.20880	.29273	-.32165
HSCOUN	.54437	-.22984	-.60437	.16479
HSRENT	.33272	.08261	.73085	.15023
HSAMEN	.50716	.22575	.44498	-.07599
HSCRWD	.44919	.58132	.20450	-.31138
HSSHAR	.24750	-.02894	.65497	.19896
HSFLAT	.36949	-.36507	-.11543	.31103
EMEACT	-.44359	.55058	-.05331	.59516
EM2ACT	-.62682	.61667	-.26921	-.03610
EMUNEM	.83773	.34922	.06469	-.00004
EMSTU	-.15816	.36427	.50922	-.39896
EMMWOM	-.65533	.01627	-.06394	.40939
EMUNYG	.69748	.22831	-.01999	-.08559
EMSEMP	-.35044	-.11005	.46716	-.30362
ETNCWP	.41208	.59493	.25177	-.33268
MMIG	.35258	.16871	.48088	.44338
MNOCAR	.94684	-.11270	-.02887	.03414
SEPROF	-.63934	-.08646	.31035	-.15151
SENMAN	-.42746	-.13158	.41560	.17031
SESMAN	-.07783	.16397	-.43342	.19521
SESS	.44229	.18022	-.23230	-.05859
SEUS	.45254	.07367	-.08405	.00968
SEEDUC	-.40759	-.12276	.58272	.04028
DYNGWK	.37293	.43919	.39637	.30987
DMIDWK	-.39963	.53143	.21276	.51208
DOLDWK	-.29592	-.40130	-.21555	-.25426
DYNPNS	.19456	-.81747	-.04802	-.32743
DOLPNS	.26572	-.73017	.16384	-.30288

A4.2 Component Loadings (cont.)

Variable	Component 5	Component 6	Component 7
DYNGCH	-.01027	-.49093	-.12892
DOLDCH	.20950	-.10250	-.25821
HHSING	-.04502	-.01499	.09424
HHCHIL	.14923	-.03696	-.14420
HHLPAR	.30841	-.01667	.11924
HXXFAM	.08669	.17547	.11339
HSOWN	-.22626	-.20719	.12283
HSCOUN	.29751	.14874	-.12486
HSRENT	-.30938	.17033	-.05253
HSAMEN	-.40257	.15412	-.14847
HSCRWD	-.00328	-.04387	.13547
HSSHAR	-.09708	.34224	-.21111
HSFLAT	.37105	-.22667	.39712
EMEACT	-.01040	.29130	.06339
EM2ACT	-.04532	.20356	.09097
EMUNEM	.14521	.05266	-.06446
EMSTU	.30884	.15282	.15259
EMMWOM	-.11754	.31641	.19824
EMUNYG	.21004	.12130	-.16571
EMSEMP	.09916	-.02844	-.14582
ETNCWP	-.10175	-.11704	.34943
MMIG	.14033	-.10444	.00350
MNOCAR	-.09276	.03553	.02323
SEPROF	.41900	-.02737	-.09437
SENMAN	.16011	-.12580	.02687
SESMAN	-.58433	.10660	-.11601
SESS	-.14493	.06247	.49573
SEUS	.06379	.24820	-.36013
SEEDUC	.37636	.01606	-.10789
DYNGWK	.10574	.30724	.21129
DMIDWK	-.09186	-.35964	-.03091
DOLDWK	.15329	.67747	.10107
DYNPNS	-.19474	-.06892	.04241
DOLPNS	-.14352	-.19693	-.00490

A4.3 9 Cluster classification - Cluster mean values.

Variable	Cluster								
	1	2	3	4	5	6	7	8	9
DYNGCH	5.45	7.65	4.31	9.14	6.77	3.64	4.24	5.26	7.14
DOLDCH	15.07	19.18	8.66	18.36	19.09	9.47	14.89	13.00	12.93
HHSING	20.84	19.60	51.06	20.59	11.16	47.13	19.58	29.55	28.71
HHCHIL	1.22	3.41	1.37	5.31	1.10	0.49	0.89	0.87	1.46
HHLPAR	4.72	13.78	11.14	5.01	3.23	18.18	3.24	6.68	7.87
HHXFAM	8.24	11.54	5.63	20.44	10.60	4.18	9.56	6.04	7.39
HSOWN	75.16	5.99	31.09	68.84	89.14	11.64	88.89	52.85	67.14
HSCOUN	14.62	88.88	7.43	7.34	5.67	79.42	3.48	32.68	5.86
HSRENT	7.58	0.83	45.74	18.64	2.41	5.42	5.14	8.34	24.23
HSAMEN	6.55	6.18	25.29	17.77	1.25	1.76	2.14	8.34	16.29
HSCRWD	0.30	0.95	1.34	4.29	0.31	0.27	0.26	0.25	0.86
HSSHAR	0.25	0.05	18.23	1.83	0.04	0.80	0.69	0.17	1.53
HSFLAT	2.78	16.47	10.80	0.79	1.83	47.22	3.63	12.72	1.90
EMEACT	63.62	64.64	70.89	63.14	73.39	49.24	60.83	59.63	63.23
EM2ACT	44.80	42.74	31.00	48.75	57.38	24.91	44.87	36.59	39.71
EMUNEM	6.36	16.06	16.37	19.79	4.75	13.09	4.33	7.13	11.30
EMSTU	3.33	2.96	6.34	6.75	4.44	2.07	6.64	2.11	4.43
EMMWOM	52.07	46.57	55.29	41.93	59.46	43.58	50.31	48.30	48.64
EMUNYG	11.68	24.09	20.66	25.37	8.69	20.56	10.57	10.83	15.83
EMSEMP	7.75	4.26	9.00	7.27	7.22	6.09	12.06	6.87	7.46
ETNCWP	4.49	5.09	10.69	46.27	4.42	5.47	4.05	3.15	13.34
MMIG	7.28	10.23	26.54	13.14	8.11	10.36	7.89	9.42	12.07
MNOCAR	38.28	60.60	58.54	58.50	18.31	72.07	22.70	50.89	55.99
SEPROF	12.47	4.42	15.57	5.76	21.94	7.42	34.29	9.36	7.73
SENMAN	15.18	7.95	28.71	9.48	20.82	10.29	23.93	13.15	16.09
SESMAN	36.43	32.77	20.80	28.96	35.63	22.36	17.33	31.23	29.74
SESS	16.79	21.97	10.91	24.03	11.34	17.73	7.27	16.38	17.24
SEUS	2.81	8.13	7.57	6.45	1.29	5.93	1.33	3.23	5.44
SEEDUC	3.89	0.78	15.14	2.64	7.16	2.60	13.95	3.06	5.03
DYNGWK	13.29	16.73	25.97	19.34	13.96	11.40	13.42	13.00	18.04
DMIDWK	24.87	22.06	29.31	26.32	32.16	16.38	23.54	22.83	26.24
DOLDWK	20.55	19.65	16.46	14.79	19.02	21.33	23.50	19.94	15.59
DYNPNS	14.05	10.31	8.63	7.94	6.66	21.02	13.39	17.23	12.86
DOLPNS	6.74	4.34	6.77	4.15	2.37	16.69	7.01	8.75	7.11

Journals: List of Abbreviations

A.A.A.G.	Annals of the Association of American Geographers
Am.Cartogr.	American Cartographer
Am.J.Epid.	American Journal of Epidemiology
Am.J.Obst.Gyn.	American Journal of Obstetrics and Gynaecology
Am.J.Sociol.	American Journal of Sociology
Am.J.Trop.Med.Hyg.	American Journal of Tropical Medicine and Hygiene
Am.Sociol.Rev.	American Sociological Review
Ann.Math.Stat.	Annals of Mathematics and Statistics
Ann.New Yk.Acad.Sci.	Annals of the New York Academy of Sciences
Aust.Comp.J.	The Australian Computer Journal
Aust.Geog.Stud.	Australian Geographical Studies
Aust.J.Soc.Iss.	Australian Journal of Social Issues
Aust.New Zeal.J.Soc.	Australian and New Zealand Journal of Sociology
Aust.Paed.J.	Australian Paediatric Journal
Brit.J.Prev.Soc.Med.	British Journal of Preventative and Social Medicine
Brit.Med.Bull.	British Medical Bulletin
Brit.Med.J.	British Medical Journal
Bull.IU.Geog.Soc.	Bulletin, Illinois Geographical Society
Can.Cart.	Canadian Cartographer
Can.Geog.	Canadian Geographer
Cart.J.	Cartographic Journal
CATMOG	Concepts and Techniques in Modern Geography
Comm.ACM.	Communications of the Association for Computer Machinery
Comm.Med.	Community Medicine
Comp.Aid.Des.	Computer Aided Design
Comp.J.	Computer Journal
East Lakes Geogr.	East Lakes Geographer
Econ.Geog.	Economic Geography
Envt.Behav.	Environment and Behaviour
Envt.Plan.	Environment and Planning
Geog.Ann.	Geografiska Annaler
Geog.Anal.	Geographical Analysis
Geog.J.	Geographical Journal
Geog.Mag.	Geographical Magazine
Geog.Rev.	Geographical Review
Hlth.Bull	Health Bulletin
Hlth.Serv.Res.	Health Services Research
Hlth.Soc.Serv.J.	Health and Social Services Journal
Hlth.Trnds.	Health Trends
Ind.Geog.J.	Indian Geographical Journal
Int.J.Hlth.Serv.	International Journal of Health Services
Int.Path.	International Pathology
Int.Yrbk.Cart.	International Yearbook of Cartography

J. Am. Med. Assoc.	Journal of the American Medical Association
J. Am. Stat. Assoc.	Journal of the American Statistical Association
J. Biosoc. Sci.	Journal of Biosocial Science
J. Chron. Dis.	Journal of Chronic Diseases
J. Epid. Com. Hlth.	Journal of Epidemiology and Community Health
J. Hlth. Soc. Behav.	Journal of Health and Social Behaviour
J. Mat. Child Hlth.	Journal of Maternal and Child Health
J. Obst. Gyn. Brit. Comm	Journal of Obstetrics and Gynaecology of the British Commonwealth
J. Pediat.	Journal of Pediatrics
J. Roy. Coll. Gen. Prac.	Journal of the Royal College of General Practitioners
J. Roy. Stat. Soc.	Journal of the Royal Statistical Society
J. Trop. Geog.	Journal of Tropical Geography
J. Trop. Med.	Journal of Tropical Medicine
Med. Care	Medical Care
Med. Care Rev.	Medical Care Review
Mill. Mem. Fund Qtly.	Millbank Memorial Fund Quarterly
New Zeal. Cart. J.	New Zealand Cartographic Journal
New Zeal. Geog.	New Zealand Geographer
New Zeal. Med. J.	New Zealand Medical Journal
Pap. Reg. Sci. Assoc.	Papers, Regional Science Association
Pop. Trnds.	Population Trends
Proc. Assoc. Am. Geog.	Proceedings of the Association of American Geographers
Prof. Geog.	Professional Geographer
Prog. Geog.	Progress in Geography
Prog. Hum. Geog.	Progress in Human Geography
Psychol. Bull.	Psychological Bulletin
Psychol. Rev.	Psychological Review
Reg. Studs.	Regional Studies
Roy. Soc. Hlth. J.	The Royal Society of Health Journal
Soc-econ. Plan. Sci.	Socio-economic Planning Sciences
Soc. Forces	Social Forces
Soc. Sci. Med.	Social Science and Medicine
Soc. Serv. Res. Gp. J.	Social Services Research Group Journal
Soc. Trnds.	Social Trends
South Afr. Geog. J.	South African Geographical Journal
Stat. News	Statistical News
Statistician	The Statistician
T. E. S. G.	Tijdschrift voor Economische en Sociale Geografie
T. I. B. G.	Transactions, Institute of British Geographers
Tech. For. Soc. Chnge.	Technological Forecasting and Social Change
Town Plan. Rev.	Town Planning Review
Urb. Ecol.	Urban Ecology
Urb. Studs.	Urban Studies

BIBLIOGRAPHY

- Achabal, D.D., Moellering, H., Osleeb, J.P. and Swain, R.W. (1978) 'Designing and evaluating a health care delivery system through the use of computer graphics', *Soc.Sci.Med.*, Vol.12D, pp1-6.
- Adelstein, A.M. (1978) 'Current vital statistics: methods and interpretation', *Brit.Med.J.*, Vol.11, pp983-987.
- Adelstein, A.M., MacDonald-Davies, I.M. and Weatherall, J. (1980) 'Perinatal Mortality: Social and Biological Factors 1975-1977', *O.P.C.S. Studies on Medical and Population Subjects*, No.41, H.M.S.O., London.
- Adelstein, A.M. and White, G.C. (1976) 'Causes of childrens' death analysed by social class' in *O.P.C.S. Studies on Medical and Population Subjects*, No.31; H.M.S.O., London.
- Aitken, G.V. and Neville, W. (1978) 'Spatial variations in New Zealand's infant mortality rate, 1970-1972', *New Zeal.Geog.*, Vol.34(1), pp41-52.
- Aitkin, M. (1980) 'A note on the selection of log-linear models', *Biometrics*, Vol.36, pp173-178.
- Akhtar, R. (1982) 'The Geography of Health - an essay and a bibliography'; Marwah Publications, New Delhi.
- Alberman, E. (1974) 'Stillbirths and neonatal mortality in England and Wales by birthweight 1953-1971', *Hlth.Trnds.*, Vol.6, pp14-17.
- Andrews, J., Davies, K.A., Chalmers, I. and Campbell, H. (1983) 'The Cardiff Births Survey: development, perinatal mortality, birthweight and length of gestation' in: P. Harper and E. Sunderland (eds.) 'Genetic Studies of the people of Wales'; University of Wales Press, Cardiff.
- Armstrong, R.W. (1966) 'Computer graphics in medical geography' in: *Proceedings of the International Geographical Union Latin American Regional Conference*, Vol.6, Special Commission on Medical Geography, Mexico, pp69-74.
- Armstrong, R.W. (1969) 'Standardised class intervals and rate computation in statistical maps of mortality', *A.A.A.G.*, Vol.59, pp382-390.
- Armstrong, R.W. (1972) 'Computers and mapping in medical geography' in: N.D. McGlashan (ed.) 'Medical Geography: Techniques and Field Studies'; Methuen, London.
- Armstrong, J.S. and Soelberg, P. (1968) 'On the interpretation of factor analysis', *Psychol.Bull.*, Vol.10, pp361-364.
- Baird, D. (1976) 'Environmental factors in relation to obstetrics', in: G.M. Howe and J.A. Loraine (eds.) 'Environmental Medicine'; William Heinemann medical books, London.

- Baird, D. and Thomson, A.M. (1969) 'The effects of obstetric and environmental factors on perinatal mortality by clinico-pathological cause' and 'Geographical differences in perinatal mortality by clinico-pathological cause', in: Butler, N.R., Alberman, E.D. and Peel, J. (eds.) 'Perinatal Problems'; Livingstone, London.
- Baird, D., Walker, J. and Thomson, A. (1954) 'The causes and prevention of stillbirths and first week deaths. Part III. A classification of deaths by clinical cause: the effect of age, parity and length of gestation on death rates by cause', *J.Obs.Gyn.Brit.Emp.*, Vol.LXI, No.4, p433-448.
- Baker, R. (1833) *Report of the Leeds Board of Health.*
- Baker, D. (1976) 'Hierarchic and non-hierarchic grouping methods', *Geog.Ann.*, Vol.58, Ser.B, pp42-58.
- Bandaranayake, R. (1983) 'Ethnic differences in disease - an epidemiological perspective; Paper presented to the Medical Geography Study Group of the IBG, July, Leeds.
- Barker, D. (1976) 'Hierarchic and non-hierarchic grouping methods: an empirical comparison of two techniques', *Geog.Ann.*, Vol.58B, pp42-58.
- Barker, D.J.P. and Rose, G. (1979) 'Epidemiology in Medical Practice'; Churchill Livingstone, Edinburgh.
- Barson, A. J. (1980) 'The contribution of pathology to perinatal audit' in: G. McIlwaine (ed.) 'Perinatal audit and surveillance; Royal College of Obstetricians and Gynaecologists, London.
- Beaumont, J.R. (1981) 'Location-allocation problems in a plane: a review of some models', *Soc-econ.Plan.Sci.*, Vol.15, pp217-229.
- Benjamin, B. (1982) 'Variation of mortality in the U.K. with special reference to immigrants and minority groups', in: D.A. Coleman (ed.) 'Demography of Immigrants and Minority Groups in the United Kingdom'; Academic Press, London.
- Bice, T., Eichorn, R. and Fox, P. (1972) 'Socioeconomic status and the use of physician services: a reconsideration', *Med.Care.*, Vol.10, pp261-271.
- Bickmore, D. and Stocking, B. (1982) 'Computer Mapping and tropical disease', *J.Trop.Med.*, Vol.85, pp19-26.
- Blakemore, J.J. and Reeve, D.E. (1977) 'CHORO - A user's guide to choropleth mapping by computer', Working Paper 3, North West Industry Research Unit, Univ. of Manchester, 82pp.
- Blunden, J.R. (1983) 'Introduction', in: N.D. McGlashan and J.R. Blunden (eds.) 'Geographical aspects of health'; Academic Press, London.
- Boyle, A.R. (1979) 'Automated cartography', *New Zeal.Cart.J.*, Vol.8(2), pp1-6.
- Bracken, M.B. (1983) (ed.) 'Perinatal epidemiology'; Oxford University Press, Oxford.

- Briggs, R. and Leonard, W.A. (1977) 'Mortality and ecological structure: a canonical approach', *Soc.Sci.Med.*, Vol.11, pp757-762.
- Brown, A.E. (1970) 'The Growth of Leicester', University of Leicester Press, Leicester.
- Brownlea, A.A. (1967) 'An urban ecology of infectious disease; City of Greater Wollongong-Shellharbour', *Aust.Geog.*, Vol.10, pp169-187.
- Brownlea, A.A. (1972) 'Modelling the Geographic Epidemiology of infectious hepatitis' in: N.D. McGlashan (ed.) *op.cit.*
- Bulmer, M.I.A. (1979) (ed.) 'Censuses, Surveys and Privacy'; Cass, London.
- Burns, C. (1967) 'Infectious hepatitis in Leicester, 1963-1966', *Brit.Med. J.*, Vol.iii, pp773-776.
- Burton, P.R. (1982) 'A geographical analysis of perinatal mortality in Leicestershire. An elective report', unpublished undergraduate thesis, Dept. of Community Health, University of Leicester.
- Butler, N.R., Alberman, E.D. and Peel, J. (1969) 'Perinatal Problems: Second Report of the 1958 British Perinatal Mortality Survey'; Livingstone, Edinburgh and London.
- Butler, N.R. and Bonham, D.G. (eds.) (1963) 'Perinatal Mortality: First Report of the 1958 British Perinatal Mortality Survey'; Livingstone, Edinburgh and London.
- CRU/OPCS/GRO(S) (1980) 'People in Britain: a census atlas'; H.M.S.O., London.
- Campbell, R., MacDonald Davies, I.M. and Macfarlane, A. (1982) 'Perinatal Mortality and place of delivery', *Pop.Trnds.*, Vol.28, pp9-12.
- Carlisle, E. (1972) 'The conceptual structure of social indicators', in: A. Schonfield and S. Shaw 'Social indicators and social policy'; Heinemann, London.
- Carter, C.O. and Peel, J. (1976) (eds.) 'Equalities and inequalities in health'; Academic Press, London.
- Cartwright, A. and O'Brien, M. (1976) 'Social class variations in health care and in the nature of general practitioner consultations' in: A. Stacey (ed.) 'The Sociology of the NHS', *Sociological Review Monograph* 22.
- Chadwick, J.G. (1961) 'Correlation between geographical distributions: a statistical technique', *Geography*, Vol.46, pp25-30.
- Chalmers, I. (1978) 'Perinatal epidemiology: strengths, limitations and possible hazards' in: R.W. Beard and S. Campbell (eds.) 'The current status of fetal heart rate monitoring and ultrasound in obstetrics'; Royal College of Obstetricians and Gynaecologists, London.
- Chalmers, I. (1979a) 'Perinatal health: the search for indices', *Lancet*, Vol.ii, pp1063-1065.

- Chalmers, I. (1979b) 'Causal thinking in perinatal medicine', *J.Mat. Child Hlth.*, Vol.4, pp188-189.
- Chalmers, I. (1980) 'The epidemiology of perinatal practice', *J.Mat. Child Hlth.*, Vol.5, pp435-437.
- Chalmers, I., Dauncey, M.E., Verrier-Jones, E.R., Dodge, J.A. and Gray, O.P. (1978) 'Respiratory distress syndrome in infants of Cardiff residents during 1965-1975', *Brit.Med.J.*, Vol.ii, pp1119-1121.
- Chalmers, I. and MacFarlane, A. (1980) 'Interpretation of perinatal statistics', in: B. Wharton (ed.) 'Topics in Perinatal Medicine'; Pitman Medical, Tunbridge Wells.
- Chalmers, I. and McIlwaine, G. (eds.) (1980) 'Perinatal audit and surveillance', Proceedings of the 8th study group of the Royal College of Obstetricians and Gynaecologists, London.
- Chamberlain, G. and Garcia, J. (1983) 'Pregnant women at work', *Lancet*, Vol.i, pp228-229.
- Chamberlain, G., Philipp, E., Howlett, B. and Masters, K. (1978) 'British Births 1970, Vol.II: Obstetric Care'; Heinemann Medical, London.
- Chamberlain, R., Chamberlain, G., Howlett, B. and Claireaux, A. (1975) 'British Births 1970, Vol.I: the first week of life'; Heinemann Medical, London.
- Chambers, I. (1978) 'Perinatal epidemiology: strengths, limitations and possible hazards', in: R.W. Beard and S. Campbell (eds.) 'The current status of fetal heart rate monitoring and ultrasound in obstetrics'; Royal College of Obstetricians and Gynaecologists, London.
- Chorley, R.J. and Haggett, P. (1965) 'Trend-surface mapping in geographical research', *T.I.B.G.*, Vol.37, pp47-67.
- Choynowski, M. (1959) 'Maps based on probabilities', *J.Am.Stat.Assoc.*, Vol.54, pp385-388.
- Claireaux, A. (1975) 'Stillbirths and first week deaths', in: R. Chamberlain et al. (eds.). op.cit.
- Clark, D., Davies, W.K.D. and Johnston, R.J. (1974) 'The application of factor analysis in human geography', *Statistician*, Vol.23, pp259-281.
- Clark, J. (1829) 'The influence of climate in the prevention and cure of chronic diseases', referred to by Gilbert (1958) op.cit.
- Clarke, J.I., Denham, C., Dewdney, J.C., Evans, I.S., Rhind, D.W. and Visvalingam, M. (1980) 'People in Britain: a census atlas'; H.M.S.O., London.
- Clarke, J.I. and Rhind, D.W. (1975) 'The relationship between the size of areal units and the characteristics of population structure', *Census Research Unit Working Paper 5*, University of Durham.
- Clarke, M. (1982) 'Perinatal audit: a tried and tested epidemiological method', *Comm.Med.*, Vol.4, pp104-107.

- Clarke, M. and Clayton, D. (1980) 'Interpreting case-control studies', in: I. Chalmers and G. McIlwaine (eds.) *op.cit.*
- Clarke, M. and Clayton, D. (1981) 'The design and interpretation of case-control studies of perinatal mortality', *Am.J.Epid.*, Vol.113(6), pp636-645.
- Clarke, M. and Clayton, D. (1983) 'Quality of obstetric care provided for Asian immigrants in Leicestershire', *Brit.Med.J.*, Vol.286, pp621-623.
- Cliff, A.D. (1970) 'Computing the spatial correspondence between geographical patterns', *T.I.B.G.*, Vol.50, pp143-154.
- Cockerham, W.C. (1978) 'Medical sociology'; Prentice-Hall, Englewood Cliffs, N.J.
- Committee of the National Birthday Trust (1978) 'Conclusions of the British Births 1970 Survey', in: G. Chamberlain et al. (eds.) *op.cit.*
- Coombes, M.G., Dixon, J.S., Goddard, J.B., Openshaw, S. and Taylor, P.J. (1983) 'Functional regions for the 1981 census', in: R.J. Johnston and D.T. Herbert (eds.), *Geography and the Urban Environment*; Wiley, London.
- Court, A. (1970) 'Map Comparisons', *Econ.Geog.*, Vol.46, pp.435-438.
- Craig, J. (1975) 'Multivariate socio-economic area classifications', *Stat. News*, Vol.28, pp28.4-28.6.
- Craig, J. (1982) 'The effectiveness of different area groupings', *Popn.Trnds.*, Vol.28, pp18-22.
- Craig, M. (1982) 'A micro-scale behavioural approach to the investigation of disease transmission in rural Bangladesh'; Paper presented to the Medical Geography Study Group of the IBG, July, Liverpool.
- Crawford, M.D., Gardner, M.J. and Sedgewick, P.A. (1972) 'Infant mortality and hardness of local water supplies', *Lancet*, Vol.1, pp988-992.
- Daly, M. S(1971) 'Characteristics of 12 clusters of wards in Greater London', *Research Report No.13*; Greater London Council, London.
- Daultrey, S. (1976) 'Principal components analysis', *CATMOG 8*, Geoabstracts, Norwich.
- Denham, C. (1980) 'The geography of the census, 1971 and 1981', *Pop. Trnds.*, Vol.19, pp6-12.
- Denham, C. and Rhind, D. (1983) 'The 1981 census and its results', in: D. Rhind (ed.) *op.cit.*
- Derning, W.E. and Stephan, F.F. (1940) 'On a least squares adjustment of a samples frequency table when the expected marginal totals are known', *Ann.Math.Stat.* Vol.11, pp427-444.
- Dewdney, J.C. (1981) 'The British Census', *CATMOG 29*, Geoabstracts, Norwich.

- Dewdney, J.C. and Rhind, D.W. (eds.) (1975) **People in Durham - a census atlas**, Dept. of Geography, University of Durham.
- Dixon, O.M. (1972) 'Methods and progress in choropleth mapping of population density', **Cart.J.**, Vol.9, pp19-29.
- Dixon, O.M. (1979) 'A map of population distribution in the Portsmouth area', **Cart.J.**, Vol.16, pp9-13.
- Donaldson, R.J. (1976) 'Urban and suburban differentials', in: C.O. Carter and J. Peel (eds.) **op.cit.**
- Donovan, J.L. (1983) 'Black people's health: a different way forward?' Paper presented to the Medical Geography Study Group of the IBG, July, Leeds.
- Douglas, J.W.B. (1948) **Maternity in Great Britain**; O.U.P., London.
- Douglas, J.W.B. (1950) 'Some factors associated with prematurity', **J.Obst. Gyn.Brit.Comm.**, Vol.57, pp143-170.
- Douglas, D. and Peucker, T. (1973) 'Algorithms for the reduction of the number of points required to represent a digitised line or its caricature', **Can.Cart.**, Vol.10, pp112-122.
- Ehrenberg, A.S.C. (1962) 'Some questions about factor analysis', **Statistician**, Vol.12, pp.191-208.
- Ehrenberg, A.S.C. (1975) **Data Reduction: Analyzing and interpreting statistical data**, John Wiley & Sons, London.
- Ellis, C. (1976) 'History in Leicester'; Recreational and Cultural Services Dept., Leicester.
- Elwood, J.H., Mackenzie, G. and Cran, G.W. (1974) 'Observations on single births to women resident in Belfast 1962-66. Part 1: Factors associated with perinatal mortality', **J.Chron.Dis.**, Vol.27, pp517-535.
- Engleman, L. (1981) 'PLR: stepwise logistic regression', in Dixon, W.J. (ed.) **BMDP Statistical Software**; University of California Press, Berkeley
- Evans, I.S. (1977) 'The selection of class intervals'. **T.I.B.G. New Series 2**, pp98-124.
- Evans, I.S. (1980) 'British Census data and their use in Quantitative geography', **Census Research Unit Working Paper**, No.17, Dept. of Geography, University of Durham.
- Evans, I.S. (1981) 'Census data handling', in: R.J. Bennett and N. Wrigley (eds.) **Quantitative Geography: a British View**; Routledge and Kegan Paul, London.
- Evans, I. S. (1983) 'Univariate analysis: presenting and summarising single variables', in: D. Rhind (ed.) **op.cit.**
- Evans, R.H. (1970) 'The expansion of Leicester in the nineteenth century', in A.E. Brown (ed.) **op.cit.**

- Evans, R.H. (1972) 'Leicester and Leicestershire 1835-1971', in: N. Pye (ed.) *op.cit.*
- Everitt, B.S. (1977) *The analysis of contingency tables*; Chapman and Hall, London.
- Eyles, J. and Woods, K.J. (1983) *The social geography of medicine and health*, Croom Helm, London.
- Ferguson, A.G. (1977) 'Probability mapping of the 1975 cholera epidemic in Kisumu district, Kenya', *J.Trop.Geog.*, Vol.44, pp53-74.
- Fingleton, B. (1981) 'Log-linear modelling of geographical contingency tables', *Envt.Plan.A.*, Vol.13, pp1539-1551.
- Fingleton, B. (1983) 'Log-linear models with dependent spatial data', *Envt.Plan.A.*, Vol.15, pp801-813.
- Forster, F. (1966) 'Use of a demographic base map for the presentation of areal data in epidemiology', *Brit.J.Prev.Soc.Med.*, Vol.20, pp156-171.
- Fox, D.J. (1972) 'Patterns of morbidity and mortality in Mexico City', *Geog.Rev.*, Vol.62, pp151-185.
- Friedman, G.D. (1980) *'Primer of Epidemiology'*; McGraw-Hill, New York.
- Garcia, J. (1982) 'Obtaining information from parents in confidential enquiries into perinatal deaths', *J.Mat.Child.Hlth.*, Vol.7, pp235-240.
- Garcia, J. and Oakley, A. (1982) 'Is early antenatal attendance so important?', *Brit.Med.J.*, Vol.i, p1474 (letter).
- Gardner, M.J. (1973) 'Using the environment to explain and predict mortality', *J.Roy.Stat.Soc.A*, Vol.36(3), pp.421-40.
- Gattrell, A. and Cole, K. (1983) *'Quantitative Geographical Explorations of small area census data'*; Research paper, Dept. of Geography, University of Salford.
- Gesler, W. (1985) *'The uses of spatial analysis in medical geography'*; Paper presented to the IBG/AAG Joint Symposium in medical geography, July, Nottingham.
- Giggs, J.A. (1973) 'The distribution of schizophrenics in Nottingham', *T.I.B.G.*, Vol.59, pp55-76.
- Giggs, J.A. (1979) 'Human health problems in urban areas', in: D.T. Herbert and D.M. Smith (eds.) *Social problems and the city*; O.U.P., Oxford.
- Giggs, J. (1982) *'Urban Health'*; Paper distributed to the Medical Geography Study Group of the I.B.G., July, Liverpool.
- Giggs, J.A. (1983) *'Ethnic status and mental illness in urban areas'*; Paper presented to the Medical Geography Study Group of the IBG, July, Leeds.

- Giggs, J.A., Bourke, J.B. and Ebdon, D.S. (1979) 'Variations in the incidence and the spatial distribution of patients with primary acute pancreatitis in Nottingham: 1969-76', *Gut*, Vol.20, pp366-371.
- Gilbert, E.W. (1958) 'Pioneer maps of health and disease in England', *Geog.J.*, Vol.124(2), p172-183.
- Girt, J.L. (1973) 'Distance to general medical practice and its effects on revealed ill-health in a rural environment'. *Can.Geog.*, Vol.17, pp154-166.
- Girt, J.L. (1978) 'A programming model of the spatial and temporal diffusion of contagious disease', *Soc.Sci.Med.*, Vol.12D, pp173-182.
- Gittus, E. (ed.) (1972) *Key variables in Social Research. Vol.1: Religion, Housing, Locality*, Heinemann Educational for the BSA, London.
- Goddard, J. and Kirby, A. (1976) 'An introduction to factor analysis', *CATMOG 7*, Geoabstracts, Norwich.
- Golding, J. (1979) 'The epidemiology of low birth-rate infants, I', *J.Mat. Child Hlth.*, Vol.4, p328.
- Golding, J. (1979) 'The epidemiology of malformation of the central nervous system', *J.Mat. Child Hlth.*, Vol.4, p328.
- Golding, J. (1979) 'The epidemiology of hypertension in pregnancy', *J.Mat. Child Hlth.*, Vol.4, p495.
- Goodman, L.A. (1970) 'The multivariate analysis of qualitative data: interactions among multiple classification', *J.Am.Stat.Assoc.*, Vol.65, pp226-256.
- Goodman, L.A. (1971) The analysis of multidimensional contingency tables: stepwise procedures and direct estimation methods for building models for multiple classification', *Technometrics*, Vol.13, pp33-61.
- Goodman, L.A. (1978) *Analysing Qualitative/Categorical data*; Abt Books, Cambridge, Mass.
- Gordon, A.D. (1981) 'Classification: methods for the exploratory analysis of multivariate data', *Monographs on Applied mobility and statistics*; Chapman and Hall, London.
- Griffiths, M. (1971) 'A Geographical Study of Mortality in an urban area', *Urb.Stud.*, Vol.8, pp111-20.
- Grimson, R.C. (1981) 'Searching for hierarchical clustering of disease: spatial patterns of sudden infant death syndrome', *Soc.Sci.Med.*, Vol.15, pp287-293.
- Gross, P. (1972) 'Economics of Health Facility Location', *Int.J.Hlth.Serv.*, Vol.2(1), pp63-84.
- Gruenewald, P. (1969) 'Stillbirth and early neonatal death', in: Butler, N.R., Alberman, E.D. and Peel, J. (eds.) *op.cit.*
- Haberman, S.J. (1978) *Analysis of Qualitative data. Volume 1: Introductory topics*; Academic Press, New York.

- Hagerstrand, T. (1967) *Innovation diffusion as a spatial process*; University of Chicago Press, Chicago.
- Haggett, P. (1972) 'Contagious Processes in a Planar graph: An Epidemiological Application', in: N.D. McGlashan (ed.), *op.cit.*
- Haggett, P. (1976) 'Hybridising alternative models of an epidemic diffusion process', *Econ.Geog.*, Vol.52, pp136-146.
- Hakim, C. (1977) 'Census based area profiles: a review', OPCS Occasional Paper, No.2, H.M.S.O.
- Hakim, C. (1978) 'Social and Community Indicators from the Census', OPCS Occasional Paper No.5, HMSO, London.
- Hall, M., Chng, P.K. and MacGillivray, I. (1980) 'Is routine antenatal care worthwhile?' *Lancet*, Vol.ii, pp78-80.
- Hart, J.T. (1971) 'The Inverse Care Law', *Lancet*, pp405-412.
- Heller, T. (1976) 'The distribution of caring facilities: a case study of East Anglia', *Development Studies Discussion Paper 12*, University of East Anglia.
- Heller, T. (1978) *Restructuring the Health Service*; Croom Helm, London.
- Hellier, J. (1977) 'Perinatal Mortality: 1950 and 1973', *Pop.Trnds*, Vol.10, pp13-15.
- Hennen, J. (1830) 'Sketches of the medical topography of the Mediterranean', referred to by Gilbert (1958) *op.cit.*
- Herbert, D.T. (1968) 'Principal Components Analysis and British studies of urban social structure', *Prof.Geog.*, Vol.20, pp280-3.
- Herbert, D. (1972) *Urban Geography: a social perspective*; David and Charles, Newton Abbot.
- Herbert, D.T. & Smith, D.M. (eds.) (1979) *Social problems and the city*, O.U.P., Oxford.
- Hinds, M.W., Bergeisen, G.H. and Allen, D.T. (1985) 'Neonatal outcome in planned v. unplanned out of hospital births in Kentucky', *J.Am.Med.Assoc.*, Vol.253, pp1578-1582.
- Hirsch, August (1883-1886) *Handbook of Geographical and Historical Pathology* (3 vols.); New Sydenham Society, London.
- Hodge, D.C. and Gatrell, A. (1976) 'Spatial constraint and the location of urban public facilities', *Env't.Plan.A*, Vol.8, pp215-230.
- Holtermann, S. (1975) 'Census indicators of urban deprivation', *Working Note No.6 'Great Britain'*, Department of Environment, ECUR Division, Mimeo.
- Hopkins, C.E., Hetherington, R.W. and Parsons, E.M. (1975) 'Quality of medical care: a factor analysis approach using medical records', *Hlth.Serv.Res.*, Vol.10, pp199-208.

- Hosmer, D.W. and Lemeshow, S. (1981) 'Goodness of fit tests for the multiple logistic regression model', referenced by Englemanop. cit.
- Howe, G.M. (1963) **National Atlas of Disease Mortality in the United Kingdom**; Royal Geographical Society, London.
- Howe, G.M. (1969) 'Computers putting diseases on the map', *Nature*, Vol.223, p891.
- Howe, G.M. (1970) 'Some aspects of social malaise in Scotland', *Hlth.Bull.*, Vol.28, ppl-23.
- Howe, G.M. (1970) 'Some recent developments in disease mapping', *Roy.Soc.Hlth.J.*, Vol.90, ppl6-20.
- Howe, G.M. (1972) **Man, Environment and Disease in Britain**; David and Charles, Newton Abbot.
- Howe, G.M. (1976) 'Aspects of medical geography', in: G.A. Harrison and J.B. Gibson (eds.) **Man in Urban Environments**, Oxford University Press, London.
- Howe, G.M. (ed.) (1977) **A World Geography of Human Diseases**; Academic Press, London.
- Howe, G.M. (1978) 'Death in London', *Geog.Mag.*, Vol.5(4), pp284-289.
- Howe, G.M. (1979) 'Statistical and cartographic methods of analysing spatial patterns of mortality', *Geoforum*, Vol.10(3), pp311-22.
- Hunter, J.M. and Young, J. (1971) 'Diffusion of influenza in England and Wales', *A.A.A.G.*, Vol.61, pp637-653.
- Imber, V. (1977a) 'A classification of the English personal social services authorities for the DHSS', **Statistical and Research Report Series No.16**, H.M.S.O.
- Imber, V. (1977b) 'Expenditure on the Personal Social Services and an area classification of need', *Soc.Serv.Res.Gp.J.*, ppl-9.
- Janson, G.G. (1969) 'Some problems of ecological factor analysis', in: M. Dogan and S. Rokkan (eds.) **Quantitative Ecological Analysis in the Social Sciences**; M.I.T. Press, Cambridge, Mass.
- Jenks, G.F. (1963) 'Generalisation in statistical mapping', *A.A.A.G.*, Vol.53, ppl5-26.
- Jenks, G.F. (1981) 'Lines, computers, and human frailties', *A.A.A.G.*, Vol.71(1), ppl-10.
- Jenks, G.F. and Caspall, F.C. (1971) 'Errors on choroplethic maps: definition, measurement, reduction', *A.A.A.G.*, Vol.61, pp217-244.
- Jenks, G.F. and Coulson, M.R.C. (1963) 'Class intervals for statistical maps', *Int.Yrbk.Cart.*, Vol.3, ppl19-33.
- Jenks, G.F. and Knos, D.S. (1961) 'The use of shading patterns in graded series', *A.A.A.G.*, Vol.51, pp316-334.

- Jiohu, R.M. (1972) 'Urban determinants of racial differentiation in infant mortality', *Demography*, Vol.9(4), pp603-616.
- Johnson, M.R.D. and Cross, M. (1983) 'Ethnic Groups and Primary Health Care in the West Midlands'; Paper presented to the Medical Geography Study Group of the I.B.G., July, Leeds.
- Johnston, R.J. (1976a) 'Area Studies, ecological studies, and social patterns in cities', *T.I.B.G.*, New Series, Vol.1, pp118-122.
- Johnston, R.J. (1976b) 'Classification in Geography', *CATMOG*, No.6, Geoabstracts, Norwich.
- Johnston, R.J. and Semple, R.K. (1984) 'Classification using information statistics', *CATMOG*, No. 37, Geoabstracts, Norwich.
- Jones, A. (1982) 'Local area mortality studies in late nineteenth century London: the cases of Kensington and Bethnal Green'; Paper presented to the Medical Geography Study Group of the I.B.G., July, Liverpool.
- Jones, K. (1982) 'The ecological fallacy: a review and alternative interpretation'; Paper presented to the Quantitative Methods Study Group (QMSG) of the I.B.G., April, Birmingham.
- Jones, K. and Kirby, A. (1980) 'The use of chi-square maps in the analysis of census data', *Geoforum*, Vol.11(4), pp409-417.
- Joseph, A.E. (1985) 'Measuring potential access to health services: some comments on the design and calibration of indices'; Paper presented to the I.B.G./A.A.G. Joint Symposium in Medical Geography, July, Nottingham.
- Joseph, A.E. and Phillips, D.R. (1984) *Accessibility and Utilization: Geographical Perspectives on health care delivery*; Harper and Row, London.
- King, P.E. (1979) 'Problems of spatial analysis in geographical epidemiology', *Soc.Sci.Med.*, Vol.13D, pp249-252.
- Kirby, A. and Scott-Samuel, A. (1981) 'Health and health care in the inner city', *Reading Geographer*, Issue 8, pp31-42.
- Kirby, A.M. and Tarn, T.D. (1976) 'Some problems of mapping the 1971 census by computer', *Envt.Plan. A*, Vol.8, pp507-513.
- Kirmeyer, S.L. (1978) 'Urban Density and pathology: a review of research', *Envt.Behav.*, Vol.10(2), pp247-270.
- Kitagawa, E.M. and Hauser, P.M. (1973) *Differential mortality in the United States: a study in socioeconomic epidemiology*, Harvard Univ. Press, Cambridge.
- Knoke, D. and Burke, P.J. (1980) *'Log-Linear Models'*; Sage Publications, Beverly Hills, London.
- Knox, E.G. (1971) 'Epidemics of rare diseases', *Brit.Med.Bull.*, Vol.27, pp43-47.

- Knox, E.G., Marshall, T., Kane, S. and Green, A. (1980) 'Social and health care determinants of area variations in perinatal mortality', *Comm.Med.*, Vol.2, p282-290.
- Knox, P.L. (1975) *Social well-being: a spatial perspective*; Oxford University Press, Oxford.
- Knox, P. (1978) 'The Intra-Urban ecology of primary medical care: patterns of accessibility and their policy implications', *Envt.Plan. A*, Vol.10, pp415-435.
- Knox, P. (1979a) 'The accessibility of of primary care to urban patients: a geographical analysis', *J.Roy.Coll.Gen.Prac.*', Vol.29, ppl60-168.
- Knox, P.L. (1979b) 'Medical Deprivation, area deprivation and public policy', *Soc.Sci.Med.*, Vol.13D, ppl11-121.
- Kohl, S.G. (1960) *'Community Obstetrical Study'*; Brooklyn, New York.
- Kohn, R. and White, K.L. (eds.) (1976) *Health care: an international study*; Oxford University Press, London.
- Lambert, P. (1976) 'Perinatal mortality: social and environmental factors', *Pop.Trnds.*, Vol.4, pp4-8.
- LAMSAC (1980) 'Specifications for a computer package for use by local authorities in analysing 1981 census small area statistics'; LAMSAC.
- Learmonth, A.T.A. (1958) 'Medical Geography in Indo-Pakistan: a study of twenty years data', *Ind.Geog.J.*, Vol.33, ppl-59.
- Learmonth, A.T.A. (1961) 'Medical geography in India and Pakistan', *Geog.J.*, Vol.127, ppl0-26.
- Learmonth, A.T.A. (1964) 'Retrospect on a project in applied regional geography in Mysore State, India', in: R.W. Steel and R.M. Prothero (eds.) *'Geographers and the tropics, Liverpool essays'*; Longmans, London.
- Learmonth, A.T.A. (1975) 'Ecological medical geography.' *Prog.Geog*, Vol.7, pp201-226.
- Learmonth, A.T.A. (1978) *'Patterns of Disease and Hunger'*; David and Charles, Newton Abbot.
- Learmonth, A.T.A. (1981) 'Geographers and health and disease studies, 1972-1980', *Soc.Sci.Med.*, Special issue, Vol.15D, pp9-19.
- Learmonth, A.T.A. and Spate, O.K. (1967) *'India and Pakistan'*; Methuen, London.
- Leete, R. and Fox, J. (1977) 'Registrar General's Social Classes. Origins and Uses', *Pop.Trnds.*, Vol.8, ppl-7.
- Leicester City Council (1984) *Survey of Leicester : Initial report of survey*; Leicester City Council and Leicestershire County Council
- Levy, L. and Herzog, A.H. (1978) 'Effects of crowding on health and social adaptation in the city of Chicago.' *Urb.Ecol.*, Vol.3(4), pp327-354.

- Lewis, G.J. (1969) 'Social Patterns in Leicester, 1670-1900', Discussion paper, University of Leicester Geography Department.
- Lewis, G.J. (1972) 'Leicester: Urban Structure and Regional relationships', in: N. Pye (ed.) *op.cit.*
- Lewis, G.J. and Davies, W.K.D. (1973) 'The urban dimensions of Leicester, T.I.B.G., Special Publication No.5, pp71-86.
- Leibenson, S. (1958) 'Ethnic groups and the practice of medicine', *Am.Soc. Rev.*, Vol.23, pp542-549.
- Lilienfeld, A.M. (1958) 'The epidemiologic method in cancer research', *J.Chron.Dis.*, Vol.8, pp647-654.
- Little, D.L. (1975) 'Social indicators and public policy', *Futures*, Vol.7, pp41-51.
- Llewellyn-Jones, D. (1974) 'Human Reproduction and Society', Chapters 15-19; Faber and Faber, London.
- Llewellyn-Jones, D. (1977) 'Fundamentals of Obstetrics and Gynaecology, Vol.1, Obstetrics'; Faber and Faber, London.
- Lloyd, R. and Steinke, T. (1977) 'Visual and statistical comparison of choropleth maps', *A.A.A.G.*, Vol.67, p429-436.
- Lovett, A.A., Bentham, C.G. and Flowerdew, R. (1985) 'Analysing Geographic Variations in mortality using Poisson Regression: the example of ischaemic heart disease in England and Wales 1969-1973'; Paper presented to the I.B.G./A.A.G. Joint Symposium in medical geography July, Nottingham.
- Lumb, K.M., Congdon, P.J. and Lealman, G.T. (1981) 'A comparative review of Asian and British-born maternity patients in Bradford, 1974-8', *J.Epid.Com.Hlth.*, Vol.35, pp106-109.
- MacDonald Davies, I.M. (1980) 'Perinatal and infant deaths: social and biological factors', *Pop.Trnds.*, Vol.19, pp19-21.
- McDowall, M., Goldblatt, P. and Fox, A.J. (1981) 'Employment during pregnancy and infant mortality', *Pop.Trnds.*, Vol.26, pp12-15.
- MacFarlane, A. (1978) 'Variations in number of births and perinatal mortality by day of week in England and Wales', *Brit.Med.J.*, Vol.ii, pp1670-3.
- MacFarlane, A. (1979a) 'Social class variations in Perinatal Mortality', *J.Mat.Child Hlth.*, Vol.4, pp337-340.
- MacFarlane, A. (1979b) 'Perinatal epidemiology: some definitions', *J.Mat. Child Hlth.*, Vol.5, pp126-129.
- MacFarlane, A. (1980) 'Some statistical approaches to comparisons between perinatal mortality rates for small areas', in: I. Chalmers and G. McIlwaine (eds.), 'Perinatal audit and surveillance'; Royal College of Obstetricians and Gynaecologists, London.

- MacFarlane, A. (1981) 'The derivation and uses of perinatal and neonatal mortality rates', *J.Pediat.*, Vol.98, pp61-62.
- MacFarlane, A., Chalmers, I. and Adelstein, A.M. (1980) 'The role of standardisation in the interpretation of perinatal mortality rates', *Hlth.Trnds.*, Vol.3(12), pp45-50.
- MacFarlane, A. and Mugford, M. (1983) 'Birth counts: statistics of pregnancy and childbirth'; H.M.S.O., London.
- McGlashan, N.D. (1966) 'The medical geographer's work', *Int.Path.*, Vol.7, pp81-83.
- McGlashan, N.D. (1972) 'Viral hepatitis in Tasmania', *Soc.Sci.Med.*, Vol.13, pp115-152.
- McGlashan, N.D. (1972) 'Medical Geography: Techniques and field studies'; Methuen, London.
- McGlashan, N. (1977) 'A note on the medical care hierarchy', *Soc.Sci.Med.*, Vol.11, pp773-774.
- McGlashan, N. (1982) 'A West Indies Geographic Pathology Survey', Occasional Paper No.12, Dept. of Geography, University of Tasmania.
- McGlashan, N.D. (1983) 'The use of cluster analysis with mortality data', in: McGlashan, N.D. and Blunden, J.R. (eds.) 'Geographical aspects of health'; Academic Press, London.
- McGlashan, N.D. and Harington, J.S. (1976) 'Some techniques for mapping mortality', *South Afr.Geog.J.*, Vol.58(1), pp18-24.
- McIlwaine, G. (1980) 'Perinatal mortality enquiries at regional level', in: I. Chalmers and G. McIlwaine (eds.) *op.cit.*
- Mackay, J.R. (1963) 'Isopleth class intervals: a consideration in their selection', *Can.Geog.*, Vol.7, pp42-45.
- McKinley, J.B. (1972) 'Some approaches and problems in the study of the use of services - an overview', *J.Hlth.Soc.Behav.*, Vol.13, pp115-152.
- McKinley, E.B. (1935) 'A Geography of Disease'; George Washington University Press, Washington.
- MacLean, I.H. (1982) 'Small area analysis of mortality patterns - some health service applications'; Paper presented to the Medical Geography Study Group of the I.B.G., July, Liverpool.
- MacVicar, J. (1980) 'Perinatal mortality - an area survey', in: I. Chalmers and G. McIlwaine (eds.) *op.cit.*
- MacVicar, J., Clarke, M., Phillips, L.F. and Hughes, A.O. (1977) 'Perinatal Mortality in Leicestershire, 1976'; University of Leicester School of Medicine.
- Maguire, D.J., Mingins, P.S., Saunders, M.N.K. and Whitelegg, J. (1983) 'Lancaster district: a computer drawn census atlas (1981)'; Dept. of Geography, University of Lancaster.

- Mallett, R. and Knox, E.G. (1979) 'Standardised perinatal mortality ratios: technique utility and interpretation', *Comm.Med.*, Vol.1(6), pp6-13.
- Mather, P.M. and Openshaw, S. (1974) 'Multivariate methods and geographical data', *Statistician*, Vol.23, pp283-308.
- Maxwell, R.J. (1981) 'Health and wealth: an international study of health care'; Lexington Books, Farnborough.
- May, J.M. (1955) 'The contribution of medical geography to the etiology of disease', *Am.J.Trop.Med.Hyg.*, Vol.4, pp776-780.
- May, J.M. (1958 & 1961) 'Studies in Medical Geography'; (2 vols.), New York.
- May, J.M. (1961) 'Studies in Disease Ecology'; Hafner Publishing Co., New York.
- May, J. (1974) 'Medical Geography: a theory of medicine', *J.Biosoc.Sci.*, Vol.6(2), pp187-192.
- May, J.M. (1977) 'Medical geography: its methods and objective', *Soc.Sci. Med.*, Vol.16, pp715-730.
- Mayer, J.D. (1982a) 'Medical Geography: some unsolved problems', *Prof.Geog.*, Vol.34(3), pp261-269.
- Mayer, J. (1982b) 'Relations between two traditions of medical geography: health systems planning and geographical epidemiology', *Prog.Hum.Geog.*, Vol.6, pp216-230.
- Mayhew, L.D. (1981) 'Automated isochrones and the locations of emergency medical services in cities', *Prof.Geog.*, Vol.33(4), pp423-428.
- Meade, M.S. (ed.) (1980) 'Conceptual & Methodological issues in Medical Geography', *Studies in Geography*, No.15, Department of Geography, University of North Carolina at Chapel Hill.
- Medical Statistics Division, O.P.C.S. (1978) 'Social and biological factors in infant mortality 1975-6', O.P.C.S. Occasional Paper, No.12, H.M.S.O.
- Mersey Region Working Party on Perinatal Mortality (1982) 'Confidential enquiry into perinatal deaths in the Mersey region', *Lancet*, Vol.1, pp491-494.
- Miller, R.D.G. and Richards, B. (1974) 'An analysis of birthweight by gestational age of infants born in England and Wales 1967-1971', *J.Obs.Gyn.Brit.Comm.*, Vol.81, pp956-967.
- Mohan, J. and Rhind, D. (1983) 'Linking and analysing census and medical data: mortality experience in Co. Durham, 1969-1977', *Census Research Unit Working Paper 21*, University of Durham.
- Mojena, R. (1977) 'Hierarchical grouping methods and stopping rules: an evaluation', *Comp.J.*, Vol.20, pp359-363.
- Mormonier, M.S. (1975) 'Class intervals to enhance the visual correlation of choroplethic maps'; *Can.Cart.*, Vol.12, pp.161-178.

- Monmonier, M.S. (1972) 'Contiguity-biased class interval selection: a method for simplifying patterns on statistical maps', *Geog.Rev.*, Vol.62, pp203-228.
- Monmonier, M.S. (1981) 'Computer-aided cartography: Principles and prospects'; Prentice-Hall, Englewood Cliffs, New Jersey.
- Morgan, C. and Denham, C. (1982) 'Census small area statistics (SAS): measuring change and spatial variation'; *Pop.Trnds.*, Vol.28, pp12-17.
- Morrill, R.L., Earickson, R.J. and Rees, P. (1970) 'Factors influencing distances travelled to hospitals'; *Econ.Geog.*, Vol.46, p161-171.
- Morris, J.N. and Heady, J.A. (1955) 'Social and Biological factors in infant mortality I'; *Lancet*, Vol.1, p343.
- Morris, J.N. and Heady, J.A. (1959) 'Social and Biological factors in infant mortality'; *Studies on Medical and Population subjects*, No.15, H.M.S.O.
- Muehrcke, P. (1973) 'Visual pattern comparison in map reading'; *Proc. Assoc.Am.Geog.*, Vol.5, pp190-194.
- Miller, G.A. (1956) 'The magical number 7, plus or minus 2? Some limits on our capacity for processing information?' *Psychol.Rev.*, Vol.63, pp81-96.
- Muller, J.C. (1975) 'Associations in choropleth map comparison'; *A.A.A.G.*, Vol.65(3), pp403-413.
- Muller, J.C. (1979) 'Perception of continuously shaded maps', *A.A.A.G.*, Vol.69(2), pp240-249.
- Muller, J.C. (1980) 'Visual comparsion of continuously shaded maps', *Cartographica*, Vol.17(1), pp40-52.
- Murray, G.D. and Cliff, A.D. (1977) 'A stochastic model for measles epidemics in a multi-region setting', *T.I.B.G., New Series*, Vol.2, pp158-174.
- Murray, M.A. (1967) 'The geography of death in the U.S. and U.K.' *A.A.A.G.*, Vol.57, pp301-304.
- Naeye, R.L., Blanc, W. and Paul, C. (1973) 'Effects of maternal nutrition on the human fetus', *Pediatrics*, Vol.52, pp494-503.
- Naeye, R.L., Kissane, J.M. and Kaufman, N. (eds.) (1981) 'Perinatal diseases'; Williams & Williams, Baltimore.
- Newcombe, R. and Chalmers, I. (1977) 'Changes in distribution of gestational age and birthweight among firstborn infants of Cardiff residents', *Brit.Med.J.*, Vol.11, pp925-926.
- Newman, D. (1978) 'Techniques for ensuring the confidentiality of census information in Great Britain', *OPCS Occasional Paper 4*, pp20-30.

- Newsholme, A. (1910) **Infant and Child Mortality**, Report C, d5263 Supplement to the 39th Annual Report of the Local Government Board, Darling and Son, Ltd., London.
- Nordstrom, S. and Beckman, L. (1981) 'Variations of birthweight in areas with different degrees of urbanisation and industrialisation.' **AMBIO**, Vol.10(1), pp24-25.
- Norris, P. (1980) 'The 1981 census: a user's perspective', **Census Research Unit Working Paper 16**, University of Durham.
- Nortman, D. (1974) 'Parental age as a factor in pregnancy outcome and child development', **Reports on Population/Family Planning**, No. 16
- Oakley, A., MacFarlane, J.A. and Chalmers, I. (1982) 'Social class, stress and reproduction', in: A.R. Rees and H. Purcell (eds.) **Disease and the environment**; John Wiley, Chichester.
- Olson, J. (1972) 'Class interval systems on maps of observed correlated distributions', **Can.Cart.**, Vol.9, pp122-131.
- Olson, J. (1972) 'The effects of class interval systems on choropleth map correlation', **Can.Cart.**, Vol.9, pp44-49.
- OPCS (1980) **Classification of Occupations**, H.M.S.O., London.
- OPCS (1979) 'Geographical area units', **Pop.Trnds.**, Vol.15, p16.
- Openshaw, S. (1977) 'A geographical solution to scale and aggregation problems in region-building, partitioning and spatial modelling', **T.I.B.G.**, New Series, Vol.2, pp459-472.
- Openshaw, S. (1982a) 'The classification of large census data sets', Paper presented at the 3rd European Colloquium on theoretical and quantitative geography, Augsburg.
- Openshaw, S. (1982b) 'Ecological fallacies and the analysis of aggregate census data', Paper presented to the Q.M.S.G. of the I.B.G., April, Birmingham.
- Openshaw, S. (1982c) 'A review of the modifiable areal unit problem in geography', Paper presented to the I.B.G., January, Southampton.
- Openshaw, S. (1982d) 'Cluster Analysis Programs'; Dept. of Geography, University of Newcastle upon Tyne.
- Openshaw, S. (1984) 'The Modifiable Areal Unit problem', **CATMOG 38**, Geoabstracts, Norwich.
- Openshaw, S. and Cullingford, D. (1979) 'Deprived places or deprived people: a study of aggregation effects inherent in area based policies', **Discussion Paper 28**, Centre for Urban and Regional Development Studies, Newcastle University.
- Openshaw, S., Cullingford, D. and Gillard, A.A. (1980) 'A critique of the national census classifications of OPCS/PRAG'. **Town Plan.Rev.**, Vol.51, pp421-439.

- Openshaw, S. and Gillard, A.A. (1976) 'An empirical study of the effects of various operational decisions on a spatial classification of census ED data'. Paper presented to IBG conference on the use and analysis of census data, University of Sheffield.
- Openshaw, S. and Gillard, A.A. (1978) 'On the stability of a spatial classification of census enumeration district data', in: P.W.J. Batey (ed.) *Theory and method in urban and regional analysis*; Pion, London.
- Openshaw, S., Sforzi, F. and Wymer, C. (1984) 'National classifications of individual and areal census data: methodology, comparisons and geographical significance'. Paper presented to I.B.G., January Conference, Durham.
- Openshaw, S. and Taylor, P.J. (1981) 'The modifiable areal unit problem', in: R.J. Bennett and N. Wrigley (eds.) *op.cit.*
- Pearson, M. (1983) 'Ethnic minority health studies: friend or foe';. Paper presented to the Medical Geography Study Group of the I.B.G., July, Leeds.
- Peckham, C.S., Ross, E.M. and Farmer, R.D.T. (1982) 'Pregnancy, childbirth and perinatal mortality', in: D.L. Miller and R.D.T. Farmer (eds.) *Epidemiology of Diseases*; Blackwell Scientific Publications, Oxford.
- Petermann, A. (1852) 'Cholera map of the British Isles showing the districts attacked in 1834, 1882, and 1833'.
- Peterson, M.P. (1979) 'An evaluation of unclassed crossed-line choropleth mapping', *Am.Cartogr.*, Vol.6 (1), pp21-37.
- Phillips, D.R. (1979) 'Spatial variations in attendance at general practitioner services'. *Soc.Sci.Med.*, Vol.13D, pp169-181.
- Phillips, D.R. (1981) 'Contemporary issues in the geography of health care'; Geo Books, Norwich.
- Powell, R.H. (1846) 'A Medical topography of Tunbridge Wells', referred to by Gilbert (1958) *op.cit.*
- Prentice, R.L. (1976) 'A generalisation of the probit and logit methods for dose response curves', *Biometrics*, Vol.32, pp761-768.
- Pritchard, R.M. (1976) 'Housing and the spatial structure of the city', Cambridge University Press, Cambridge.
- Provin, R.W. (1977) 'The perception of numerosness on dot maps', *Am.Cartogr.*, Vol.4(2), pp111-125.
- Pugh, J. (1962) 'Urban Geography of Leicester'; Unpublished University of Leicester thesis (M.A.).
- Purola, T., Kalimo, K., Sievers, K. and Nyman, K. (1968) 'The Utilisation of the Medical Services and its relationship to Morbidity, Health Resources and social factors'; Research Institute for Social Security, Helsinki.
- Pye, N. (ed.) (1972) *Leicester and its Region*; University of Leicester Press, Leicester.

- Pyle, G. (1969) 'Diffusion of cholera in the United States', *Geog.Anal.*, Vol.1, pp59-75.
- Pyle, G.F. (1976) 'Foundations to Medical Geography', *Econ.Geog.*, Vol.52, pp95-102.
- Pyle, G.F. (1979) *Applied Medical Geography*; Wiley, New York.
- Rathwell, T. (1984) 'General Practice, ethnicity and health services delivery', *Soc.Sci.Med.*, Vol.19 (2), pp123-130.
- Ratowsky, D.A. and Lance, G.N. (1978) 'A criterion for determining the number of groups in a classification', *Aust.Comp.J.*, Vol.10, pp115-117.
- Redfern, P. (1980) 'Census 1981 - an historical and international perspective', *Pop.Trnds.*, Vol.23, pp3-15.
- Rees, P.H. (1971) 'Factorial Ecology: an extended definition, survey and critique', *Econ.Geog.*, Vol.47, pp220-233.
- Rhind, D.W. (1974) 'High speed maps by laser beam', *Geog.Mag.*, Vol.46 (8), pp393-394.
- Rhind, D.W. (1975) 'Geographical analysis and mapping of the 1971 U.K. Census data'. Census Research Unit Working Paper No.3. Dept. of Geography, University of Durham.
- Rhind, D.W. (1977) 'Computer Aided Cartography', *T.I.B.G.*, New Series 2 Vol.1, pp71-96.
- Rhind, D.W. (1981) 'Geographical information systems in Britain', in: R.J. Bennett and N. Wrigley (eds.), *op.cit.*
- Rhind, D.W. (ed.) (1983) *A Census User's Handbook*, Methuen, London.
- Rhind, D.W., Evans, I.S. and Dewdney, J.C. (1977) 'The derivation of new variables from population census data', *Census Research Unit Working Paper*, No.9, University of Durham.
- Rhind, D.W., Evans, I.S. and Visvalingam, M. (1980) 'Making a national atlas of population by computer', *Cart.J.*, Vol.17, pp3-11.
- Rhind, D.W., Rogers, A. and Dawson, J. (1979) 'Why digitize?' *Area*, Vol.11 (3), pp211-214.
- Roberts, I.M.L. (1969) 'The Census and Research: Ideals and Realities', *T.I.B.G.*, Vol.48, pp173-187.
- Robinson, A.H. (1956) 'The necessity of weighting values in correlation analysis of areal data', *A.A.A.G.*, Vol.46, pp233-236
- Robinson, A.H. and Bryson, R.A. (1957) 'A method for describing quantitatively the correspondence of geographical distributions', *A.A.A.G.*, Vol.47, pp379-391.
- Robinson, A.H., Sale, R. and Morrison, J. (1978) *Elements of Cartography*; John Wiley, New York.

- Robinson, G.K. (1950) 'Ecological correlation and the behaviour of individuals', *Am.Soc.Rev.*, Vol.15, pp351-357.
- Roemer, M.I. (1977) 'Systems of health care'; Springer, New York.
- Roff, A. (1977) 'Census data, normality and transformations'; Urban and Regional Studies Unit, University of Kent.
- Roff, A. (1977) 'The importance of being normal', *Area*, Vol.9, pp195-198.
- Rooth, G. and Bratterby, L.-E. (eds.) (1976) 'Perinatal Medicine'; Almquist and Wiksell International, Stockholm.
- Rose, G.A. (1976) 'Epidemiological evidence for the effects of urban environments', in: G.A. Harrison and J.B. Gibson (eds.), *Man in Urban Environments*; Oxford University Press, London.
- Salkever, D.S. (1976) 'Accessibility and the demand for preventive care', *Soc.Sci.Med.*, Vol.10, pp469-475.
- Salmond, G.F. (1974) 'Social needs for medical services: the inverse care law in New Zealand', *New Zeal.Med.J.*, Vol.80, pp396-403.
- Schoeman, M.E.F. and Mahajan, V. (1977) 'Using the Delphi method to assess community health needs', *Tech.For.Soc.Chnge*, Vol.10 (2), pp203-210.
- Scott-Samuel, A. (1977) 'Social area analysis in Community Medicine', *Brit.J.Prev.Soc.Med.*, Vol.31, pp199-204.
- Scripter, M.W. (1970) 'Nested means map classes for statistical maps', *A.A.A.G.*, Vol.60, pp.385-393.
- Shafer, S. (1980) 'Mapping bone cancer death rates in Pennsylvania counties', *Soc.Sci.Med.*, Vol.14D, pp11-15.
- Shah, F.W. and Abbey, H. (1971) 'Effects of some factors on neonatal and postneonatal mortality', *Mill.Mem.Fund.Qtly.*, Part I, Vol.49, pp33-57.
- Shannon, G.W., Bashshur, R.L. and Metzner, C. (1969) 'The concept of distance as a factor in accessibility and utilization of health care', *Med. Care Rev.*, Vol.26, pp143-161.
- Shannon, G.W. and Dever, G.E.A. (1974) *Health Care Delivery: spatial perspectives*, McGraw-Hill, New York.
- Shapter, T. (1849) *History of the Cholera in Exeter in 1832*.
- Shepherd, I.D.H. (1979) 'Harvard computer graphics Conference Report', *Cart.J.*, Vol.16 (1), pp40-43.
- Skevky, E. and Bell, W. (1955) *Social Area Analysis*; Stanford University Press.
- Shortridge, B.G. (1979) 'Map reader discrimination of lettering size', *Am.Cartogr.*, Vol.6 (1), pp13-20.
- Sillitoe, K. (1978a) 'Ethnic Origins 1: an experiment in the use of a direct question about ethnicity for the census', *OPCS Occasional Paper*, No.8, H.M.S.O.

- Sillitoe, K. (1978b) 'Ethnic origin: the search for a question', *Pop.Trnds*, Vol.13, pp25-30.
- Sills, A., Tarpey, M.R. and Golding, P. (1981) *Inner Area Reserach Project Social Survey 1st Report*; University of Leicester Centre for Mass Communications Research.
- Silk, J. (1979) *Statistical Concepts in Geography*; George, Allen and Unwin, London.
- Simmons, J. (1972) 'Leicester past and present', in: A.E. Brown (ed.) *op.cit.*
- Simmons, J. (1974) *Leicester Past and Present. Vol.1: Ancient Borough to 1860*; Eyre Methuen, London.
- Simmons, J., Whayne, T.F., Anderson, G.W. and Horack, H.M. (1944-1954) *Global epidemiology*, (3 vols.), J.B. Lippincott, Philadelphia.
- Sixsmith, A.J. (1982) 'The spatial planning of hospital facilities'; Paper presented to the Medical Geography Study Group of the I.B.G., July, Liverpool.
- Slysz, W.D. (1974) 'An evaluation of statistical software in the social sciences', *Comm.A.C.M.*, Vol.17, pp3-6.
- Smith, D.M. (1974) 'Who gets what, where and how; a welfare focus for human geography', *Geography*, Vol.59, pp289-297.
- Smith, D.M. (1977) *Human geography: a welfare appraoch*; Edward Arnold, London.
- Smith, D.M. (1979) *Where the Grass is Greener: Living in an Unequal World*; Penguin, Harmondsworth.
- Snow, J. (1855) *On the mode of communication of cholera*; 2nd edition, London.
- Social Services Committee (1980) *Perinatal and neonatal mortality. 2nd Report from the SSC 1979-80*; H.M.S.O., London.
- Spicer, C.C. and Lipwroth, L. (1966) 'Regional and social factors in infant mortality', *Studies on Medical and Population Subjects*, No.19, H.M.S.O.
- Sprague, H.A. and May, J.J. (1978) 'Family age in the epidemiology of family health', *Soc.Sci.Med.*, Vol.12, pp271-275.
- Steward, H.J. (1974) 'Cartographic generalisation: some concepts and explanation', *Cartographica Monograph*, No.10.
- Stimson, R.J. (1983) 'Research design and methodological problems in the geography of health' in: N. McGlashan and J. Blunden (eds.) *op.cit.*
- Stock, R. (1976) *Cholera in Africa*; International African Institute, London.

- Stockwell, E.G. (1962) 'Infant Mortality and Socio-Economic Status: A changing relation', *Mill.Mem. Fund Qtly.*, Vol.40, pp101-111.
- Stockwell, E.G. and Adamchak, D.J. (1978) 'Changing patterns of infant mortality among social areas in Toledo, Ohio 1950-70', *East Lakes Geogr.*, Vol.13, pp62-68.
- Strachan, A.J. (1972) 'Patterns of Population Change', in: N. Pye (ed.), *op.cit.*
- Struening, E.L., Rabkin, J.G., Cohen, P., Raabe, G., Muhlin, G.L. and Cohen, J. (1973) 'Family, ethnic and economic indicators of low birth weight and infant mortality: a social area analysis', *Ann. New Yk. Acad.Sci.*, Vol.218, pp87-107.
- Suchman, E.A. (1964) 'Sociomedical variations among ethnic groups', *Am.J.Sociol.*, Vol.70, pp319-331.
- Summer, G. (1971) 'Trends in the location of primary medical care in Britain', *Antipode*, Vol.3, pp46-53.
- Taylor, D.R.F. (1979) 'Development atlases using computer techniques', *Cartographica Monograph*, No.23, pp67-78.
- Taylor, D.R.F. (1980) *The computer in contemporary cartography*; Wiley, New York.
- Taylor, P.J. (1977) *Quantitative methods in Geography*; Houghton Mifflin Co., Boston.
- Taylor, S.M., Myers, J., White, N. and Frank, J. (1985) 'Modelling the incidence of childhood diarrhoea'; Paper presented to the I.B.G./A.A.G. Joint Symposium in Medical Geography, July, Nottingham.
- Thomas, C.J. (1976) 'Sociospatial differentiation and the use of services', in: D.T. Herbert and R.J. Johnston (eds.) *Social Areas in Cities*; Wiley, London.
- Thompson, J.F. (1968) 'Some Observations on the Geographic Distribution of Premature Births and Perinatal Deaths in Indiana', *Am.J.Obst.Gyn.*, Vol.101, p43-52.
- Tobler, W.R. (1965) 'Computation of the correspondence of geographical patterns', *Pap.Reg.Sci.Assoc*, Vol.15, pp131-139.
- Tobler, W. (1973) 'Choropleth maps without class intervals?' *Geog.Anal.*, Vol.3, pp262-265.
- Townsend, E. (1969) 'Problems in designated areas', *Lancet*, Vol.1, pp1179-1190.
- Townsend, P. (1979) *Poverty in the United Kingdom*; Penguin, Harmondsworth.
- Townsend, P. and Davidson, N. (1982) *Inequalities in Health*; Penguin, Harmondsworth.
- Tukey, J.W. (1977) *Exploratory Data Analysis*; Addison-Wesley, Reading, Massachusetts.

- U.M.R.C.C. (1982) '1981 census data at UMRCC', University of Manchester.
- Unwin, D.J. (1981) *Introductory Spatial Analysis*: Methuen, London and New York.
- Upton, G.J.G. (1978) *The analysis of cross-tabulated data*; Wiley, Chichester.
- Upton, G.J.G. (1981) 'Log-linear models, screening and regional industrial surveys', *Reg.Studs.*, Vol.15, pp33-45.
- Upton, G.J.G. and Fingleton, B. (1979) 'Log linear models in geography';, *T.I.B.G.*, New Series, Vol.4 (1), pp103-115.
- U.S. Bureau of the Census and Manpower Administration (1974) *Urban Atlas*; U.S. Government Printing Office, Washington, D.C.
- Visvalingam, M. (1976) 'Chi-square as an alternative to ratios for statistical mapping and analysis', *Census Research Unit Working Paper*, No.8, Dept. of Geography, University of Durham.
- Visvalingam, M. (1979) 'The signed chi-square measure for mapping', *Cart.J.*, Vol.16, pp93-98.
- Visvalingam, M. (1981) 'Area based social indicators: problems and prospects', *Census Research Unit Working Paper* 20, University of Durham.
- Visvalingam, M. and Dewdney, J.C. (1977) 'The effects of the size of areal units on ratio and chi-square mapping', *Census Research Unit Working Paper*, No.10, Dept. of Geography, University of Durham.
- Wachs, M., and Kumagai, T.G. (1973) 'Physical accessibility as a social indicator', *Soc.-econ.Plan.Sci.*, Vol.7, pp437-456.
- Wallis, S.M. and Harvey, D.R. (1980) 'Small for dates babies, their problems and their future', in: B. Wharton (ed.) *op.cit.*
- Walker, J. (1980) 'Classification', in: I. Chalmers and G. McIlwaine (eds.) *op.cit.*
- Walmsley, D.J. (1978) 'The influence of distance on hospital usage in rural New South Wales', *Aust.J.Soc.Iss.*, Vol.13, pp72-81.
- Walters, V. (1980) *Class inequalities and Health Care*; Croom Helm, London.
- Weatherall, J.A.C. (1975) 'Infant Mortality: International differences', *Pop.Trnds*, Vol.1, H.M.S.O.
- Webber, R.J. (1975) 'Liverpool Social Area Study 1971 data: final report', *PRAG Technical Paper*, No.14, Centre for Environmental Studies, London.
- Webber, R. (1977a) 'The social structure of Haringey', *PRAG Technical Paper*, No.18, Centre for Environmental Studies, London.
- Webber, R.J. (1977b) 'Multivariate Analysis of Small Area Census Data: a review of PRAG experience', *PRAG Technical Paper* No.19, Centre for Environmental Studies, London.

- Webber, R. (1977c) 'Cumbria social area analysis', PRAG Technical Paper No.22, Centre for Environmental Studies, London.
- Webber, R. (1979) 'Census enumeration Districts: a socio-economic classification', OPCS Occasional Paper, No.14, H.M.S.O.
- Webber, R.J. and Craig, J. (1978) 'A socio-economic classification of Local Authorities in Great Britain', OPCS Studies in Medical and Population Subjects, No.35, H.M.S.O.
- Weil, C. and Weil, J. (1980) 'Time allocation research and health problems', in: M.S. Meade (ed.) *op.cit.*
- Weinstein, M. S. (1980) 'Health in the City; Pergamon Press, New York.
- Weiss, J. and Greenlick, M. (1970) 'Determinants of Medical Care Utilisation', *Med.Care*, Vol.8, pp456-462.
- Whitelegg, J. (1982) *Inequalities in Health: Problems of Access and Provision*; Straw Barnes Press, Retford.
- Wharton, B.A., Smalley, C., Millns, C., Nirmal, J., Bissenden, J.G. and Scott, P.H. (1980) 'The Asian mother and her baby at Sorrento', in: B. Wharton (ed.) *op.cit.*
- W.H.O. (1978) 'Report on the social and biological effects on perinatal mortality'; W.H.O. Geneva, Switzerland.
- Wigglesworth, J. (1980) 'Monitoring perinatal mortality - a pathophysiological approach', *Lancet*, Vol.ii, pp684-686.
- Willie, C.V. (1959) 'A research note on the changing association between infant mortality and socio-economic status', *Soc.For.*, Vol.37, pp221-227.
- Wilson, M.G.A. (1973) 'A note on infant death in Melbourne', *Aust.Paed.J.*, Vol.8, pp61-71.
- Wilson, M.G.A. (1978) 'The geographical analysis of small area/population death rates: a methodological problem', *Aust.Geog.Stud.*, Vol.16 (2), pp149-160.
- Wishart, D. (1978) 'CLUSTAN User Manual (3rd edition)'; Program Library Unit, Edinburgh University.
- Wood, M. (1968) 'Visual perception and map design', *Cart.J.*, Vol.5, pp54-64.
- Wood, J. and Barr, R. (1982) 'Inequalities in Urban Primary Care'; Paper presented to the Medical Geography Study Group of the I.B.G., July, Liverpool.
- Wrigley, N. (1976) 'An introduction to the use of logit models in geography', *CATMOG*, No.10, Geoabstracts, Norwich.
- Wrigley, N. (1985) 'Categorical data analysis for geographers and environmental scientists'; Longman, London.

- Wrigley, N. and Bennett, R.J. (eds.) (1981) **Quantitative Geography: A British view**; Routledge and Kegan Paul, London.
- Yankauer, A. (1950) 'The relationship of fetal and infant mortality to residential segregation', *Am.Soc.Rev.*, Vol.15, pp644-648.
- Yerushalmy, J. (1956) 'Longitudinal studies in pregnancy', *Am.J.Obst.Gyn*, Vol.71, pp80-96.
- Young, J.C. (1977) 'Bronchitis', in: G.M. Howe (ed.) **op.cit.**