
Methodological issues in cross-border analyses of European small-area data: a case study

Sally Cook, Michael A Poole, Adrian J Moore

School of Environmental Studies, University of Ulster, Coleraine, Northern Ireland;
e-mails: s.cook@ulst.ac.uk; ma.poole@ulst.ac.uk; a.moore@ulst.ac.uk

Dennis G Pringle

Department of Geography, National University of Ireland, Maynooth, County Kildare, Republic of Ireland; e-mail: dennis.pringle@may.ie

Received 17 May 1999; in revised form 8 September 1999

Abstract. The growing trends for political and economic integration within the European Union signal a need for improved cohesion of national data sets to allow for comparative cross-national research on socioeconomic topics. Difficulties in achieving comparability of variables and spatial bases can apply whether analysis is undertaken at the most macro level (that of entire nations) or at more micro spatial scales. In this paper we will discuss the methodological problems which can arise from comparative cross-border analysis by using small-area census data within Ireland (North and South) as illustration. As this analysis demonstrates, limited overlap between the specific variables in national data sets, different national contexts for socioeconomic indicators, and variations between data sets in the size of spatial units can all cause analytical problems at this spatial scale. In particular, scale mixing can potentially result in very misleading interpretations. Possible solutions to these problems are discussed, and policy recommendations are made in relation to microscale spatial data sets at the national, as well as at the pan-European, level.

1 Introduction

There is a tendency among geographers and other social scientists when studying small-area population characteristics to be strongly constrained by political and national boundaries. This leads to their concentrating on demographic and socio-economic variations within, but not beyond, the borders of an individual country. To a large extent this may be explained and even justified by the motivation for undertaking studies of this nature, particularly when it rests upon an attempt to mould or monitor national policies, as, for example, in the case of much spatial analysis of social deprivation.

On the other hand, if we widen the spatial perspective, cross-national comparison can yield useful insights across the spectrum of socioeconomic conditions. Indeed the increasing impetus towards harmonisation of national and regional population statistics, at least within the European Union (EU) (Tanenbaum and Mochmann, 1994; Waters, 1995), testifies to the perceived value of research which reaches beyond the bounds of a single country. Such research may employ one-off survey data or depend on routinely collected statistics, covering topics as diverse as living standards among the elderly (Tsakoglou, 1996), changes in unemployment and inequality (Blank, 1995), patterns of part-time employment (Fagan and Rubery, 1996), and relationships between education and social class (Ishida et al, 1995). The importance of cross-national data to social science is, according to Jowell (1998, page 168), “incontestable”.

However, comparative European research remains relatively uncommon, despite an abundance of national data (Tanenbaum and Mochmann, 1994), and above all there is a distinct paucity of studies which employ small-area population data in a cross-national context. Surely, if single-nation small-area studies and cross-national regional-level research are both considered to be valid and useful tools for spatial analysis, then the

use of small-area data for analysis of demographic or socioeconomic conditions in a cross-border context could potentially be valuable too. Perhaps the narrow focus on a single jurisdiction is understandable for state-level policymakers responsible for national decisionmaking and answerable to national electorates. However, it must often be a less satisfactory option for academics in the research community, many of whom would prefer to take a wider view of the world and to be less constrained by the existence of political boundaries which are only one factor amongst a myriad causes of spatial variability.

Much of this neglect of cross-border studies may in fact be explained simply by the practical difficulties of conducting such research by using data which are collected and released by different national administrations which operate independently. This independence means that data may lack sufficient mutual compatibility to deliver meaningful results for effective comparative research. European data, even at national and regional level, are not well integrated (Tananbaum and Mochmann, 1994) and, in fact, the problem of true equivalence of data is a "recurring theme in international comparative research" (Hakim, 1991, page 110). This leads, according to Jowell (1998), to a tendency to accept data as comparable when the questions and techniques employed are not functionally equivalent, the end result being that national differences observed may simply be artifactual. Indeed there is no doubt that such practical difficulties alone present a significant, and perhaps at times insurmountable, barrier to cross-national small-area research.

It is this issue of data comparability which forms the focus of the present paper, with the example of Ireland to illustrate some of the methodological problems encountered in the analysis of cross-border small-area data. The obvious source of information at this scale level in both parts of Ireland is the population census, for both the Irish Republic and the United Kingdom are amongst that majority of European states which still undertake a regular conventional census (Langevin et al, 1992). We will focus, for illustrative purposes, on the use of census data to identify areas of deprivation, but similar principles apply to other uses of this source of small-area data.

Two specific themes are investigated. The first involves problems relating to the choice of variables for analysis, including not only national differences in the definition, operationalisation, and categorisation of variables, but also, in a more subtle sense, inconsistencies, emanating from different national contexts, in the real meaning of what are superficially similar variables. The second theme is more specifically geographical and concerns the actual spatial units for which small-area data are made available. It addresses both national differences in the size of these areas and also scale variation *within* national data sets. In our opinion, it is this issue, involving scale mixing in the data, which constitutes the major problem to be addressed.

Before focusing on these two methodological themes, however, we need to begin by setting the local Irish scene for their exploration. This initially involves reference to the accelerating trend towards cross-border cooperation on the island, a trend which is stimulating the need for cross-border analysis. It is then necessary to identify Irish data sources, for it is their specific limitations that provide the problem that we seek to address in this paper.

The Irish context to this research is important because of the recommendations for local policy change, in relation to the modification of data collection policy on the two sides of the Irish border, which have resulted and which have already been put forward informally to the authorities in both jurisdictions. However, there is a much wider application to these policy recommendations, for it is inconceivable that the Irish frontier is the only one in Europe across which cross-border small-area analysis is either appropriate or encounters problems. Therefore, highlighting the methodological

issues which arise when using Irish data for cross-national analysis could further inform debate on matters pertaining to cross-national research in general. This is particularly important in the context of an EU ever more conscious of the need to allocate resources in a spatially sensitive way that is transparently legitimate.

2 The Irish policy context for cross-border research

Many geographically detailed studies of socioeconomic conditions have been undertaken within both Northern Ireland (Boal et al, 1974; 1978; Goodyear and Eastwood, 1978; Project Team, 1976; Robson et al, 1994) and the Republic of Ireland (Haase, 1995; 1999; Howell et al, 1993; Nolan et al, 1998; 1999; SAHRU, 1997; Williams, 1993). However, these studies, often motivated by the targeting element of area-based policies, have all been carried out in national isolation, without reference to small-area variations on the opposite side of the border. Although this is not particularly surprising in view of the fact that at present such area-based policies are self-contained within each state, it is ever more likely, with the development of the Irish peace process and the prospect of cross-border bodies with executive powers, that they will have an increasingly cross-border emphasis in years to come.

There is already a growing tendency for cooperation between state agencies at an all-Ireland level, in areas such as tourism, roads, energy, and health (O'Dowd et al, 1995). Irish cross-border cooperation has also been encouraged both by the Anglo-Irish Agreement and by EU policies and programmes, especially the EU structural funds (such as INTERREG) which have been developed to promote social and economic cohesion. Recent examples of cross-border cooperation in the social sphere include the Ballyconnell agreement between Health Boards adjoining the border (which examined the possibility of providing integrated health-care services), the BORDER project which aims to provide a wide-ranging Web-based spatial information service with direct relevance to social and economic issues, and a number of local community-based job creation schemes, funded, for example, by the EU's Peace and Reconciliation Programme and the International Fund for Ireland. Particularly important in the context of this paper is the current speculation that small-area data will in future be employed in the Irish border region to target, monitor, and evaluate EU INTERREG funding. This will obviously necessitate the use of a cross-border data source which can produce fully comparable information to enable fair and unbiased area-based targeting.

The move towards cross-border cooperation received a major boost with the signing of the Good Friday Agreement at Stormont on 10 April 1998, subsequently ratified by the majority of the electorate in separate referenda held in Northern Ireland and the Irish Republic in May 1998. The Good Friday Agreement makes provision for a North-South Ministerial Council to promote cooperation between the two jurisdictions. Those attending the Ministerial Council will have decisionmaking powers, whilst remaining accountable to the Northern Ireland Assembly and the Oireachtas (Irish Parliament). Although progress towards these objectives has been slower than some may have anticipated, the long-term trend is undoubtedly towards closer cooperation and a harmonisation of government structures.

3 Socioeconomic data sources in Ireland

A wide-ranging socioeconomic analysis requires a diversity of variables, which are available from only two types of source. One of these is the sample survey. However, in practice, samples designed for national surveys do not generate sufficiently reliable findings to permit statements about geographical areas smaller than counties or even groups of counties. On the other hand, locally intensive sampling cannot be done on a

sufficiently extensive scale to permit valid statements about more than a handful of hopefully representative small areas. That leaves the population census as the only source with the 100% coverage required to allow reliable results to be derived about a comprehensive network of small geographical areas. Inevitably, on grounds of both financial cost and ease of understanding, the census suffers from the limitation of a simple questionnaire, so the resulting variables are less ambitious in scope than those from a customised sample survey. However, the 100% coverage is its crucial advantage, and discussion of the methodological problems of cross-border small-area analysis necessarily focuses on this source.

Cross-national analysis of spatially referenced data must be undertaken with regard to three separate dimensions of the data, namely variable type, area base, and time. The simplest of these three issues is that of time, and it can be dealt with here without having to be further elaborated upon later in the paper. It is unfortunate that, whereas the Irish Republic normally takes its census quinquennially, Northern Ireland follows the United Kingdom practice of being only decennial. The result is that, for half of each decade, comparability in time between the two censuses in Ireland can be achieved only by ignoring the most recent census south of the border and by using an older one instead. For instance, at the time of writing, the results of the 1996 census have been published for the Irish Republic, but its 1991 census results have to be used for any comparison with Northern Ireland because that is the year of the most recent census north of the border. Fortunately, in every other respect, the two censuses are synchronised, with both jurisdictions having had a census on exactly the same day in April 1991.

The actual seriousness of the implications of the Northern Ireland census being only decennial depends on the rate of socioeconomic change taking place either nationally or locally. It has been particularly unfortunate in the 1990s that cross-border comparison in Ireland has had to continue to employ the 1991 baseline. This is because the flourishing economy of the Irish Republic, usually labelled the 'Celtic Tiger' effect, has been primarily a post-1991 feature (Sweeney, 1998) and cannot be reflected in a cross-national census comparison with Northern Ireland until 2001.

Such a macro-level economic effect has, of course, micro-level implications, though it is well known, and indeed a cause for concern, that not all small localities have benefited equally from this Celtic feline (Duggan, 1999). In addition, any intercensal period in any part of the world will have some neighbourhoods experiencing socio-economic change. However, it is rare for neighbourhood transformation to be so profound that the change through time is as radical as the cross-sectional variation through space (Poole, 1995). As a result, patterns of spatial variation tend to be remarkably stable, particularly in the span of a single decade.

The other two issues, of variables and area base, are more complex and, in our opinion, much more serious because, while the selection of indicator variables and area bases for a single-state analysis is restricted by the limitations of one national data set, cross-national analysis is constrained further by the degree of overlap between two separate data sets. It is the exploration of these two issues which forms the core of this paper, dealing with problems which impact both on attempts to produce cross-national geodemographic classifications (Waters, 1995) and on studies even at regional and national level (Tanenbaum and Mochmann, 1994). They are also analogous to those which affect international survey research (Jowell, 1998).

In Ireland the census has been undertaken by two different authorities since the first postpartition censuses of 1926, and although the data collected for Northern Ireland differ only marginally from that of Great Britain, for instance with respect to ethnicity and religion, there has been little apparent coordination of data collection or

release between the administrations in Dublin and Belfast (Compton, 1993; Linehan, 1991–92). Some cross-border similarity survives by inertia from the common prepartition origin—for example the use of religion—but substantial differences in the two data sets, which affect both the comparability of indicator variables and the spatial scale of data, have evolved.

4 Specification of variables

The degree of divergence between the two Irish census data sets is manifested immediately by the fact that approximately 9000 variables, in 75 cross-tabulations, are available from Northern Ireland's small-area statistics, whereas the Republic's data consist of only 1750 counts. Such differences in data release can give rise to problems of nonoverlap in terms of both the general domain of socioeconomic variable, and the specific definition of an indicator.

If we deal first with variable domain, figure 1 illustrates the theoretical overlap of some of the more commonly applied social indicators which are released for each jurisdiction. Although several important types of data are held in common, it can also be seen that certain variables are not reproducible for both territories. Despite the greater overall detail present in the Northern Irish data—for instance, in relation to household crowding—some key topics covered by the Southern census are not available for the North. One example is the number of persons involved in small-scale farming activity, which is crucial in an Irish context because of the underemployment and poverty which are associated with economic dependence on small farms (Jackson and Haase, 1996). The census statistics for the Republic differentiate farming activity on the basis of the acreage farmed, whereas the UK census classifies farmers only as employers/managers or as working on their own account. No further detail can be acquired from the Northern Irish data on this topic, leaving no possibility of deriving a comparable indicator of the potential deprivation due to agricultural underemployment.

Turning to specifics, there are many examples of variables with overlapping dimensions for which specific definitions are either restricted or do not match at all. Data on economic activity, for example, are released for different age bands in the two data sets; age categories in the Republic's data are coarser than the North's, starting

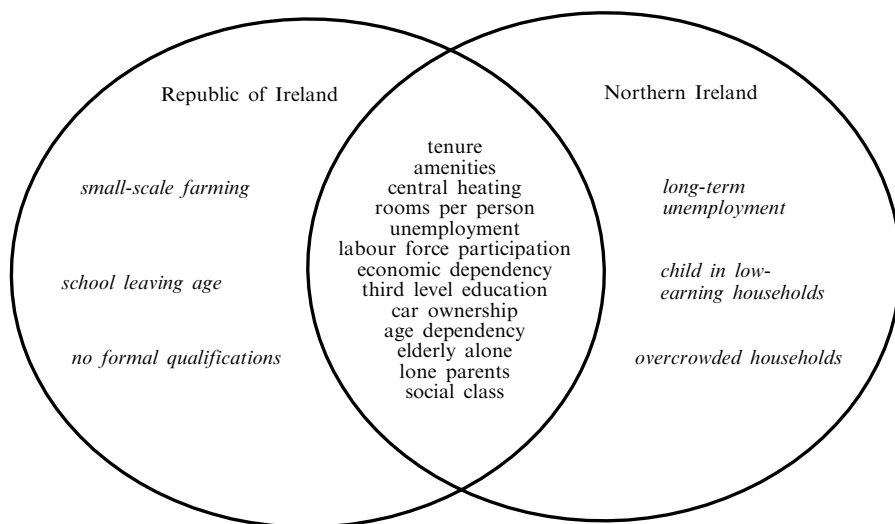


Figure 1. Overlap of socioeconomic data domains in census small-area statistics for Northern Ireland and the Irish Republic.

at age fifteen and being divided into ten-year intervals, thus crossing the normal age of female retirement. In relation to some of the other variables, it is the Irish Republic which has the stronger data. For example, its census output includes eleven categories of educational attainment ranging from no formal qualifications to postgraduate degrees. In Northern Ireland, on the other hand, though the census form actually categorised seven different qualification levels from none to 'degree level or higher', only third-level qualifications are identified in the small-area statistics. Therefore, Irish cross-border research cannot differentiate people with no qualifications from those with low or moderate attainment.

A third issue involving variables is that of national contexts. This refers to variables which are comparable in terms of actual definition, but which have overall national rates which differ so markedly between each country that in terms of their socio-economic interpretation they cannot be considered to be equivalent. A particularly important example is that of housing tenure, with local authority tenancy in particular being used as a social indicator on both sides of the border. Living in local authority housing is associated with a risk of poverty which is well above average in the Republic of Ireland (Nolan et al, 1998) and the United Kingdom (Davies et al, 1997). It is probably fair to suggest that, in both jurisdictions in Ireland, the public housing sector increasingly has a residual function in supplying housing to families who would have difficulty affording to buy their own property (Burrows, 1999). Therefore, in each state, the public sector largely houses the most deprived segment of the population. It follows that mapping the proportion of all housing in the public sector in each small geographical area is a very useful way of indicating where the major spatial concentrations of deprivation are located *within each state*. However, the overall proportion of people in public housing in Northern Ireland is much higher than in the Republic of Ireland—29% as opposed to 10%—and this is not because the Northern jurisdiction is vastly more deprived than its Southern neighbour. Instead it reflects past differences in government policy: in the South investment in public-sector housing in urban areas was more limited, whereas in rural areas the higher proportion of farmers, with their owner-occupier status, had the same effect (O'Connell, 1994). Consequently, public housing tends to cater only for the extremely deprived of the South, whereas it serves both this group and the moderately deprived in the North. The result is that a cross-border analysis with this variable would identify both moderate and extreme deprivation in Northern Ireland but would miss many areas that were equally deprived in the Republic of Ireland.

5 Areal units

It has long been known that as a result of the modifiable areal unit problem, the scale at which spatial data are analysed profoundly influences the outcome of that analysis, so that results at one scale level are far from necessarily replicated at another (Amrhein, 1995; Openshaw, 1984). There follows a consequence, however, which, despite being stressed by Wright (1937) over sixty years ago, is not always quite so well appreciated. This is that a single piece of spatial analysis, or a single tier in a multilevel design, needs to adopt a constant scale level. The reason is that, if spatial results are scale-specific, then they must relate to a single specific scale level and not to a set of geographical subareas which are so disparate in size that they have to be described as involving scale mixing. If we accept that, in social science, scale is normally conceived in terms of the basic population elements in the analysis, it follows that having constant scale means using geographical subareas of approximately equal population.

An attempt to identify socioeconomic characteristics among small communities in Ireland by using census statistics necessarily means undertaking analysis based

on areas defined for administrative purposes, rather than their appropriateness for academic or even policy-oriented study. The smallest such areas for which data are made available are the district electoral division (DED) in the Republic of Ireland, and the enumeration district (ED) in the North. However, the application of the label 'small-area' to both these sets of spatial units does not necessarily imply equivalence in terms of spatial scale, either internally within the two jurisdictions or externally in a comparison between them. Like administrative areas all over the world, they fail to have constant population size and so are subject to some degree of scale mixing.

Comparison of the distribution of population size for EDs and DEDs shows that the average population of small areas in the Republic is more than double that of those in the North (at 1024 compared with 423). However, the median figures for the two countries are far closer, at 490 and 426, respectively. Clearly, therefore, the frequency distribution is very skewed in the Republic, where the mean is inflated by a relatively small number of outliers with excessively large populations, mainly in urban areas. The largest Southern DED population is, at 25 843, more than ten times that of the largest Northern ED. Of course, the enormous variation in scale which is contained within the DED data causes a problem which is not unique to cross-border analysis, having serious repercussions for analysis within the Irish Republic itself. However, use of combined North–South data yields a particular set of problems for analysts because of the systematic national bias which is introduced when making comparisons such as rank-order listings.

The most extreme outliers amongst the Republic's small areas are found not in the five main cities, which have county borough status and are subdivided into wards (with an average population well under 3000), but in some of the other large towns. The majority of settlements in the Republic are composed of only one or two DEDs. The result is that a town of 5000 people in the Republic can be expected to have a DED population ten times as large as the average urban ED in Northern Ireland. Extreme examples are provided by the towns of Drogheda and Dundalk (two of the three largest towns in the Republic outside of the county boroughs), both of which for 1991 census data release consisted of only one urban DED, with minor urban overspill into adjoining DEDs. The populations of the undivided urban units for these two towns are 23 848 and 25 843, respectively. These may be contrasted with several towns in Northern Ireland with total populations in the size range 20 000–30 000, such as Ballymena, Lurgan, Portadown, Newtownards, and Newry, each of which are comprised of more than forty EDs. In other words, the spatial mesh of the small-area statistics in Drogheda and Dundalk is more than forty times coarser than in these Northern Ireland towns.

Lurgan is the Northern Irish town closest in size to Dundalk, with an overall male unemployment rate of 23% in 1991. Yet, although this town-level unemployment figure is 6% lower than that of Dundalk's, a third of Lurgan's EDs have rates which exceed the rate in Dundalk's urban DED, the highest being 59%. Whereas this Lurgan ED achieves a rank of 66 out of the entire Irish dataset of 7173 EDs and DEDs, Dundalk ranks only 1274. Even the Northern town of Newtownards, which had an overall unemployment rate of just 11%, contains an ED which, at 938, ranks considerably higher than that of Dundalk. It is not possible to obtain quantitative information on the level of sociospatial segregation within Dundalk, but there is little doubt that parts of the town, along with a number of other towns in the Republic, would rank far higher were data to be released for smaller, more homogeneous, zones. Small-scale variation inside some towns is therefore effectively being concealed by the form of census data release.

The outcome of unadjusted comparative urban analysis incorporating small-area data as published for the two states is therefore that the more homogeneous spatial

units in Northern towns tend to dominate strongly the extreme ends of joint rank-order listings. Taking again the example of male unemployment, only four out of the twenty highest ranked small areas in Ireland, and eight out of the highest fifty, are located south of the border. Such empirical evidence is not conclusive, but it is certainly suggestive, for there is nothing in the relative unemployment figures, north and south, at more macro levels to suggest that this social problem was much greater in Northern Ireland than in the Republic in 1991. Therefore the concentration of the worst small areas for unemployment in the North seems more likely to be a statistical artefact, resulting from the problem of scale mixing, rather than a real situation on the ground.

Significantly, the majority of the highest ranked areas are located in the larger towns and cities, where strong residential segregation at a meso-scale level leads to the unemployed being concentrated in small pockets well revealed in most of the small-area data. This is exactly the same pattern that was revealed by the research of Moore (1995, page 122) into the geography of the Jarman Index for measuring deprivation in a health-care context. It is therefore particularly pertinent to test the persistence of the Northern bias in the data by switching attention to settlements of much smaller size. Thus, if villages and small towns (in the population range 1500–5000) are chosen, the Northern EDs are still found to dominate the extremes—only one of the first twenty and five out of the highest fifty areas are located in the Republic of Ireland (table 1). This is because these places are still large enough to comprise several EDs in Northern Ireland—a town of 5000 people will on average have ten to twelve EDs—but they will

Table 1. Highest ranked small areas for male unemployment, based on joint Northern Ireland–Republic of Ireland data for settlements of population range 1500–5000: areas in the Republic are in italics. (Northern Ireland small-area statistics are Crown Copyright 1994, reproduced by permission of the Controller of The Stationery Office. Source: 1991 Northern Ireland Census, Crown Copyright, ESRC/JISC/DENI purchase; Small Area Population Statistics for the Republic of Ireland are copyright of the Central Statistics Office, Dublin).

Rank	Area name, census code, and county or DCA	ED/DED population	Male unemployment (%)
1	Crossmaglen 0905, Newry and Mourne	370	68.1
2	Crossmaglen 0903, Newry and Mourne	331	59.0
3	Keady 1104, Armagh	517	54.0
4	Coalisland South 1102, Dungannon	673	53.1
5	Coalisland South 1104, Dungannon	447	52.6
6	Crossmaglen 0906, Newry and Mourne	350	52.2
7	Coalisland South 1103, Dungannon	569	50.4
8	Castleberg 0203, Strabane	648	50.0
9	<i>Glentogher 079, Donegal</i>	<i>684</i>	<i>50.0</i>
10	Knocklayd 1402, Moyle	532	49.7
11	Derrymore 1103, Newry and Mourne	577	48.8
12	Irvinestown 1403, Fermanagh	389	47.6
13	Lisnaskea 1802, Fermanagh	682	47.5
14	Lisnaskea 1802, Fermanagh	447	45.1
15	Dungiven 0503, Limavady	800	44.4
16	Upper Glenshane 1504, Limavady	701	43.7
17	Keady 1105, Armagh	473	43.6
18	Enagh 1703, Derry	533	43.4
19	Annalong 0104, Newry and Mourne	549	42.8
20	Crossmaglen 0904, Newry and Mourne	373	42.7

DCA, district council area; DED, district electoral division; ED, enumeration district.

usually consist of a single DED in the Irish Republic. This confirms the dominance of the bias resulting from the frequency of large so-called 'small areas' in the Republic.

6 The search for comparability: recommendations

There are then two major and distinct types of problem confronting the aspiring cross-border analyst in Ireland. One involves the choice of variables, the other the choice of geographical areas. In relation to the first, particular care must be taken not only that comparability of variable definitions is ensured, but that the impact of national differences, such as provision of public housing, upon the meaning of socioeconomic indicators does not invalidate their use. These restrictions obviously have repercussions for the breadth of possible indicators available for analysis.

It is not proposed to take the discussion of this issue any further, however, because the general problem of the choice of comparably defined variables is one which has received significant attention already in the social science literature on comparative analysis. It has been necessary in this paper to investigate the specific manifestation of this problem in the context of small-area analysis, but the solution to the difficulties identified has much in common with the general problem of the need for comparably constituted variables. In the short term, it involves the restriction of comparison to variables sharing both the same domain and similar specific definitions in the separate jurisdictions being compared, with safeguards to ensure a real similarity of fundamental meaning. In the longer term, it involves efforts to increase the harmonisation of these aspects of data across jurisdictions. Such harmonisation is perhaps politically sensitive—not least in Ireland—and also, without duplication of presentation, runs the risk of creating data discontinuities over time. This is of course always a problem when proposals for change are made in data presentation.

This same distinction—between short-term and long-term solutions—can be made in relation to the second problem discussed in this paper, concerning the choice of geographical areas. However, this issue has received so little attention in the social science literature that a more extensive discussion of the solutions to the problem is necessary, with a view to developing recommendations for both analysis and policy.

There is a major bias inherent in unadjusted analysis of cross-national data which embodies the degree of scale mixing present in Irish census output. Valid options for circumventing this problem in the short term are highly limited because the data are not available in a form which enables larger areas to be disaggregated into a finer mesh. Therefore, to avoid the totally spurious results generated by scale mixing, we consider that there are only two realistic alternatives. In terms of the terminology introduced by Haggett (1965, pages 205–210) to deal with what, following Wright (1937), he called the problem of irregular collecting areas, these are elimination and aggregation. The first of these involves omitting the larger outlier units, such as Dundalk and Drogheda in this Irish case, from the analysis altogether, and the second involves aggregating released data for small units up to the level of the larger units. Both of these short-term palliatives are, however, beset with limitations.

Of these two alternatives, omission of the larger outliers substantially defeats the purpose of the whole exercise. This is because, if the objective is to identify, say, the most deprived geographical places in the study region, outlier omission replaces this objective with one of identifying only those deprived places for which satisfactory data are available. Instructively, however, the net effect is exactly the same as conducting an unadjusted analysis of the data for all the original areas in the data set, because small pockets of deprivation within relatively large heterogeneous units like Dundalk and Drogheda cannot be identified. In fact, the only substantial difference is that omission of outliers is much more honest. An unadjusted analysis leaves the impression that

everywhere has been included, which is true, and that everywhere has a chance of having its deprivation identified if it is severe enough, which is certainly not true.

On the other hand, the second alternative, involving aggregation of all microdata up to the level of the coarsest elements in the spatial mesh, also defeats the purpose of the research. Again, if the objective is to identify small pockets of severe deprivation, then upward aggregation destroys the entire rationale for using micro-level data. In these circumstances, the analyst would have been better to use coarse-mesh data in the first place, especially as almost certainly a greater variety of variables would have been available at this more macro level. The size of the largest undisaggregated 'small area' in Ireland is close to 26 000 people. Therefore, one strategy behind aggregation to a constant scale level would be to devise a national matrix of geographical areas with a population of about this size. However, such a spatial mesh would indiscriminately cover cities, towns, villages, and countryside, and would vary in composition from swathes of rural space including small towns, villages, and open countryside to urban neighbourhoods which formed only parts of cities such as Dublin and Belfast.

More meaningful perhaps would be an alternative strategy of recognising the significance of settlement structure, and undertaking the upward aggregation to the level of the individual town. This would produce a constant scale analysis in the sense that each urban element in the spatial mesh was a single settlement, though of course it would be of variable scale in the sense that population would vary enormously from one unit to another.

Whether the upward aggregation is to a mesh of constant population or to a set of single settlements, the problem of comparing fairly homogeneous areas, like the EDs of Northern Ireland towns, with highly heterogeneous DEDs such as Dundalk and Drogheda is effectively removed by making all spatial units heterogeneous. This may be seen really as an exercise in damage limitation, and results in a lamentable loss of spatial detail, but the important point is that it nonetheless achieves an unbiased cross-border analysis by ensuring that like is compared with like. This is empirically confirmed by the application of upward aggregation to the level of the individual Irish settlement, for there emerges a more even representation of Northern and Southern settlements amongst those ranked highest on unemployment. For example, at this spatial level nine out of the highest twenty unemployment rates for villages and small towns are located in the Republic, leaving the remaining eleven in Northern Ireland (table 2). This is much less imbalanced than the results of the unadjusted data analysis (table 1) and is plausible in view of the similar overall levels of unemployment on the two sides of the Irish border. The advantage of this method over that of aggregation to units of constant population size is that relatively deprived villages, for example, will still be recognisable, whereas under a strictly constant scale level these villages could potentially be combined with other settlements and intervening countryside with very different socioeconomic characteristics, thus disguising their deprivation. This distinction would be particularly significant if the purpose of analysis embraced the targeting of resources towards deprived locations outside of large urban areas as well as inside them.

Whichever method of aggregation was chosen in an attempt to nullify the inherent bias produced by the scale mixing of the data, subsequent research involving urban areas necessarily loses its small-area status. However, given the consequences of scale mixing, valid options are very limited, for it is an unfortunate but unavoidable fact that use of the raw data intact to undertake small-area analysis means giving false credence to the label 'small area', which could result in serious misinterpretation, thereby invalidating the cross-border dimension to the research.

Table 2. Highest ranked towns for male unemployment, based on joint Northern Ireland – Republic of Ireland data for entire settlements of population range 1500 – 5000: areas in the Republic are in italics. (Northern Ireland small-area statistics are Crown Copyright 1994, reproduced by permission of the Controller of The Stationery Office. Source: 1991 Northern Ireland Census, Crown Copyright, ESRC/JISC/DENI purchase; Small Area Population Statistics for the Republic of Ireland are copyright of the Central Statistics Office, Dublin).

Rank	Town	Town population	Male unemployment (%)
1	Crossmaglen, Newry and Mourne	1424	54.6
2	Coalisland, Dungannon	3852	45.4
3	<i>Carndonagh, Donegal</i>	2565	38.4
4	Keady, Armagh	2402	37.6
5	Dungiven, Limavady	2739	37.1
6	Newtownstewart, Strabane	1484	36.8
7	Castleberg, Strabane	2100	35.9
8	Lisnaskea, Fermanagh	2221	33.7
9	Irvinestown, Fermanagh	1938	33.7
10	<i>Mountmellick, Laoighis</i>	2495	31.7
11	<i>Gorey, Wexford</i>	4429	30.4
12	Strathfoyle, Derry	1650	30.3
13	Sion Mills, Strabane	1178	30.2
14	Ballycastle, Moyle	3664	30.2
15	<i>Kilrush, Clare</i>	2740	30.1
16	<i>Rathkeale, Limerick County</i>	1803	29.6
17	<i>Edenderry, Offaly</i>	3525	29.0
18	<i>Muine Bheag, Carlow</i>	2573	28.8
19	<i>Duleek, Meath</i>	2496	28.7
20	<i>Castleblayney, Monaghan</i>	2029	28.6

There are other possible solutions to the problem of scale mixing, as was long ago made clear by Haggett (1965), including what are now called polygon overlay and other forms of spatial interpolation (Lam, 1983; Martin, 1991). Moreover, there is an attraction in certain inferential techniques, like signed χ^2 and Bayesian estimation, which have been applied to the ranking of places of disparate size in which the ordering is interpreted as being affected by random influences (Langford, 1994; Visvalingam, 1978). However, in considering these options elsewhere, we have had to conclude that they are all even less adequate as a solution to the scale-mixing problem than aggregation or elimination.

The unsatisfactory nature of all these short-term palliatives means that a radical long-term solution is particularly important for tackling the problem of the size of geographical areas. Substantial improvements have been made on both sides of the Irish border in recent decades in the release of small-area data. Indeed, specifically on the issue of the spatial matrix, 1991 witnessed a substantial improvement in Northern Ireland as, under the impetus of ESRC, the practice of standard ED data release was extended to the province from the British mainland for the first time (Caven et al, 1998). This radically augmented both the much coarser matrix of ward data and the extremely expensive, albeit superbly detailed, availability of grid-square information. However, the administrative areas for which small-area data are released south of the border continued to suffer the major technical problem of scale mixing which the Northern EDs are largely free of, and this actually prohibits the undertaking of uniformly small-area analysis. Dundalk and Drogheda are examples we have quoted,

but other towns too are inadvertently discriminated against because of the excessively large geographical areas for which their data are released.

The result is that any public policy designed to identify small geographical areas, meriting positive discriminatory treatment because of their deprivation, is never going to be recognised in Dundalk and a number of other medium-sized towns in the Republic of Ireland by means of census data because of an accident of their census geography. It is therefore, in our opinion, absolutely essential that the census authorities in the Irish Republic give further thought to the matrix of geographical areas for which they release census data, in order to eliminate these extraordinary anomalies. The main need of course is to subdivide the larger DEDs into much smaller spatial units.

7 The broader European context

This is not specifically an Irish problem, however. It is a matter for concern wherever it is intended to undertake cross-border small-area analysis. In this context, it is encouraging that, within the EU, Eurostat has recognised the importance of geographical scale in identifying a five-tier hierarchy of geographical areas within each EU state. This is the so-called NUTS system, with each tier differentiated from those above and below on the basis of spatial scale (Waters, 1995). The most micro level of these tiers, NUTS-5 (which includes areas such as DEDs in the Irish Republic, *communes* in France, and *municipios* in Spain), is the obvious one for cross-border small-area comparison within the EU.

However, close examination of this tier reveals that the average size of the spatial units identified varies widely from one country to another. Thus, in four states, the NUTS-5 areas have average populations below 2500, with the Irish ones the smallest at 1024. However, at the other extreme, in four others the average population exceeds 16 000—rising to 30 026 in Sweden (Eurostat, 1995; 1996). It is much less encouraging that, having recognised scale differentiation as fundamental, Eurostat is prevented by the strength of decisionmaking power at the national level from ensuring that there is actual scale consistency across international boundaries at any one NUTS level. Yet even this inconsistency does not embrace the scale-mixing problem that we have identified within the Irish data. This is because the national averages—subject to considerable international variation though they be—mask much greater variation at the NUTS-5 level within individual states. For example, though the Irish Republic's average is the smallest of the fifteen EU states, its largest NUTS-5 areas of Dundalk and Drogheda have almost as large a population as the national average in the EU state with the highest mean, Sweden.

Moreover, this problem of scale mixing within individual states is by no means confined to the Irish Republic. In France, the NUTS-5 level is the commune, described by Pinchemel (1987, pages 194–196, 258–260) as not only of ancient origin but still playing a role of fundamental importance in both the socioeconomic affairs and the emotional loyalties of French people. There are 36 000 of these communes, averaging about 1500 people each, but their most outstanding feature is the scale mixing to which they are subject. At one extreme, there were 103 communes with over 50 000 inhabitants in 1990 (the largest being Paris, with over 2 million), and, at the other, 1000 communes contained fewer than 50 people. Indeed, there were 28 000 communes with less than 1000 inhabitants (Arnaud, 1991, pages 14–16). There seems, in fact, to be much more uniformity amongst these communes in terms of the area of land occupied than in terms of the number of people contained.

Because of the impact of such scale mixing on the ability to identify socioeconomic problems in truly small areas, there seems to be an urgent need to explore this issue on

an international basis. Thus, although contributions such as that of Waters (1995) identifying the mean population and area within the states of the EU are invaluable, the work needs to be extended to cover the variation in size too. After all, it is a statistical truism to assert that the average conveys very little meaning if unsupported by information on dispersion. After measuring the extent of scale mixing in the purportedly small-area data of EU states, the next step is then to urge national statistical agencies to refine their data presentation to eliminate the problem. Our analysis demonstrates that scale mixing is certainly a problem which undermines the validity of cross-border analysis in Ireland, and the French material suggests that its impact may not be unimportant on the European mainland too.

8 Conclusion

Despite the enhanced awareness of the European dimension which has come with the increased integration of institutions within the enlarged EU, nationally partitioned mind-sets are still arguably the norm in the context of socioeconomic data analysis. Indeed, international boundaries are the social science equivalent of geological unconformities when it comes to considering information. The ease with which so many things—people and their goods and money, rivers, rainfall, and wildlife, and even information itself—can move across EU borders only serves to emphasise the anomaly that data incompatibilities impede international comparison, even within a small island like Ireland, let alone within a multinational region like the EU.

This means that there is an 'obvious need for data harmonisation' (Tanenbaum and Mochmann, 1994), and indeed efforts are going on to develop pan-European systems to characterise socioeconomic conditions within small areas. The geodemographic Euro-MOSAIC classification system covers seven different European countries, providing consistent information, albeit by using different area bases (Waters, 1995). The EU-funded Euripides project aimed to provide a wide variety of demographic and economic statistics for small administrative areas across Europe (at the EU NUTS-5 level), for use by the private sector, public authorities, and academic researchers, although the end product was not in fact released. In a similar vein, Eurostat is in the process of compiling a database (called SIRE) of population data, again at NUTS-5 level, to include around thirty variables.

Development of this type of system is certainly an important step in the right direction, although it would obviously be desirable if this could eventually be extended to embrace a much broader range of variables, including for example vital statistics and measures of disposable income. Significantly, however, the emphasis is on the design of variables which are to be harmonised across international boundaries. This is part of a long tradition of international statistical organisations warning users of the potential hazards of cross-border data and urging them to inspect the fine print and take careful cognisance of precise definitions when interpreting data. Such warnings refer particularly to the definition of variables, and this reinforces our view that our observations in this paper on the first of the two major problems discussed is in line with a well-established practice of urging caution to data consumers making international comparisons.

On the other hand, the second problem, dealing with the spatial units, has traditionally received considerably less attention both in the Irish context examined in this paper and, more importantly, at the European level. For instance, a report produced by the United Nation's Economic Commission for Europe and the Statistical Office of the European Communities (UN/ECE, 1997) making recommendations for population and housing censuses includes no discussion of spatial units, restricting recommendations to those concerning variables. There seems to be an astonishing tendency to be

captivated by the labels of geographical areas, whether these be the DEDs of the Irish Republic or the communes of France or any other NUTS-5 level of spatial unit. Analysts frequently seem to imagine that they are employing spatially consistent data if their spatial units all have the same name. Thus, within a single country, there is a temptation to believe that scale has been standardised if all the units have a similar label. Correspondingly, analysts making cross-national comparisons may believe that they are comparing like with like provided that they work at a single NUTS level, such as NUTS-5 at the most micro level.

What we have attempted to emphasise in this paper is that data users must exercise a similar caution when examining their spatial units as they do when inspecting the precise definition and meaning of their variables. As the data presented by Waters (1995) show, the potential impact of scale mixing in cross-national comparison, even considering only national average figures, is considerable; yet on top of this, as our Irish research has demonstrated, there will be additional scale mixing within each of these countries at NUTS-5 level, which could add dramatically to the magnitude of overall scale mixing within a cross-national study.

It really is imperative that potential users of such cross-European data sets be aware of the serious implications of data inconsistencies on their work when they undertake comparative analysis and take all steps feasible to prevent possible misinterpretation caused by these inconsistencies. Equally, in relation to the longer term, it is vital that national and international statistical agencies take steps both to eliminate serious spatial scale mixing in national data sets, and to harmonise more closely at the international level. However, whereas to achieve the standardisation of variables the overwhelming priority is international harmonisation—for definitions are usually consistent within national boundaries—the emphasis is different in relation to scale standardisation. This is because, though there are differences between national average sizes of spatial units at any one NUTS level, major discrepancies also occur *within* national data sets, as our analysis of NUTS-5 in Ireland has demonstrated. This, we believe, is a major policy issue for national statistical agencies whose significance currently seems to be underappreciated.

Acknowledgements. The authors would like to thank Professor Brian Graham for commenting on an earlier draft of the paper.

References

- Amrhein C G, 1995, "Searching for the elusive aggregation effect: evidence from statistical simulations" *Environment and Planning A* **27** 105–119
- Arnaud R, 1991 *La France en Chiffres 1992* (Hatier, Paris)
- Blank R M, 1995, "Changes in inequality and unemployment over the 1980s—comparative cross-national responses" *Journal of Population Economics* **8**(1) 1–21
- Boal F W, Doherty P, Pringle D G, 1974 *The Spatial Distribution of Some Social Problems in the Belfast Urban Area* (Northern Ireland Community Relations Commission, Belfast)
- Boal F W, Doherty P, Pringle D G, 1978, "Social problems in the Belfast urban area: an exploratory analysis", OP-12, Department of Geography, Queen Mary College, London
- Burrows R, 1999, "Residential mobility and residualisation in social housing in England" *Journal of Social Policy* **28**(1) 27–52
- Caven N, Evans T, Beatty R, 1998, "The Northern Ireland census of population and housing: looking back and looking forward", paper presented to the Statistical and Social Inquiry Society of Ireland, Belfast; copy available from the authors
- Compton P A, 1993, "Population censuses in Northern Ireland: 1926–1991", in *The 1991 Census User's Guide* Eds A Dale, C Marsh (HMSO, London) pp 330–351
- Davies H, Joshi H, Clarke L, 1997, "Is it cash that the deprived are short of?" *Journal of the Royal Statistical Society A* **160**(1) 107–126
- Duggan C, 1999, "Locally based interventions to combat poverty and exclusion: how effective can they be?" *Administration* **47**(2) 56–77

- Eurostat, 1995 *Regions: Nomenclature of Territorial Units for Statistics* (Office for Official Publications of the European Community, Luxembourg)
- Eurostat, 1996 *Population, Households and Dwellings in Europe* (Office for Official Publications of the European Community, Luxembourg)
- Fagan C, Rubery J, 1996, "The salience of the part-time divide in the European Union" *European Sociological Review* **12** 227–250
- Goodyear P M, Eastwood D A, 1978, "Spatial variations in level of living in Northern Ireland" *Irish Geography* **11** 54–67
- Haase T, 1995, "The designation of disadvantaged areas in the local development programme", report to Area Development Management Ltd to facilitate the designation of areas under the Operational Programme for Local Urban and Rural Development, 1995–1999; copy available from the authors
- Haase T, 1999, Affluence and deprivation: a spatial analysis based on the 1991 Census of Population", in *Poor People, Poor Places* Eds D G Pringle, J Walsh, M Hennessy (Oak Tree Press, Dublin) pp 13–35
- Haggett P, 1965 *Locational Analysis in Human Geography* (Edward Arnold, London)
- Hakim C, 1991, "Cross-national comparative research on the European Community: the EC Labour Force Surveys" *Work, Employment and Society* **5**(1) 101–117
- Howell F, O'Mahoney M, Devlin J, O'Reilly O, Buttanshaw C, 1993, "A geographical distribution of mortality and deprivation" *Irish Medical Journal* **86** 96–99
- Ishida H, Muller W, Ridge J M, 1995, "Class origin, class destination, and education: a cross-national-study of 10 Industrial Nations" *American Journal of Sociology* **101**(1) 145–193
- Jackson J A, Haase T, 1996, "Demography and the distribution of deprivation in rural Ireland", in *Poverty in Rural Ireland* Eds C Curtin, T Haase, H Tovey (Combat Poverty Agency, Dublin) pp 59–85
- Jowell R, 1998, "How comparative is comparative research?" *American Behavioral Scientist* **42** 168–177
- Lam N S-N, 1983, "Spatial interpolation methods: a review" *American Cartographer* **10** 129–149
- Langevin B, Begeot F, Pearce D, 1992, "Censuses in the European Community" *Population Trends* **68** 33–36
- Langford I H, 1994, "Using empirical Bayes estimates in the geographical analysis of disease risk" *Area* **26** 142–149
- Linehan T P, 1991–92, "History and development of Irish population censuses" *Journal of the Statistical and Social Inquiry Society of Ireland* **26**(4) 91–125
- Martin D, 1991 *Geographic Information Systems and their Socioeconomic Applications* (Routledge, London)
- Moore A J, 1995, "Deprivation payments in general practice: some spatial issues in resource allocation in the UK" *Health and Place* **1**(2) 121–125
- Nolan B, Whelan C T, Williams J, 1998 *Where Are Poor Households?* (Oak Tree Press/Combat Poverty Agency, Dublin)
- Nolan B, Whelan C T, Williams J, 1999, "Spatial aspects of poverty and deprivation in Ireland", in *Poor People, Poor Places* Eds D G Pringle, J Walsh, M Hennessy (Oak Tree Press, Dublin) pp 37–74
- O'Connell C, 1994, "Housing in the Republic of Ireland: review of trends and recent policy measures" *Administration* **42** 159–169
- O'Dowd L, Corrigan J, Moore T, 1995, "Borders, national sovereignty and European integration: the British-Irish case" *International Journal of Urban and Regional Research* **19** 272–285
- Openshaw S, 1984 *The Modifiable Areal Unit Problem* Concepts and Techniques in Modern Geography series, number 38 (Geo Books, Norwich)
- Pinchemel P, 1987 *France: A Geographical, Social and Economic Survey* (Cambridge University Press, Cambridge)
- Poole M A, 1995, "The spatial distribution of political violence in Northern Ireland: an update to 1993", in *Terrorism's Laboratory: The Case of Northern Ireland* Ed. A O'Day (Dartmouth, Aldershot, Hants) pp 27–45
- Project Team, 1976 *Belfast Areas of Special Social Need* (HMSO, London)
- Robson B, Bradford M, Deas I, 1994 *Relative Deprivation in Northern Ireland* Centre for Urban Policy Studies, Manchester University, Manchester
- SAHRU, 1997, "A national deprivation index for health and health services research", technical report number 2, Small Area Health Research Unit, Department of Community Health and General Practice, Trinity College, Dublin

-
- Sweeney P, 1998 *The Celtic Tiger* (Oak Tree Press, Dublin)
- Tanenbaum E, Mochmann E, 1994, "Integrating the European database—infrastructure services and the need for integration" *International Social Science Journal* **46** 499–511
- Tsakoglou P, 1996, "Elderly and non-elderly in the European Union: a comparison of living standards" *Review of Income and Wealth* **42** 271–291
- UN/ECE, 1997, "Recommendations for the 2000 Censuses of Population and Housing in the ECE Region", Statistical Standards and Studies, number 49 (United Nations, Geneva)
- Visvalingam M, 1978, "The signed chi-square measure for mapping" *Cartographic Journal* **15** 93–98
- Waters R, 1995, "Data sources and their availability for business users across Europe", in *GIS for Business and Service Planning* Eds P Longley, G Clarke (GeoInformation International, Cambridge) pp 33–47
- Williams J, 1993, "Spatial variations in deprivation surrogates—a preliminary analysis", report to the Combat Poverty Agency, Economic and Social Research Institute, Dublin
- Wright J K, 1937, "Some measures of distributions" *Annals of the Association of American Geographers* **27** 177–211