eurostat

DG 34/E

Document E/LOC/20

Luxembourg, 2.4.1992

EUROSTAT

Regional Statistics and Accounts

STUDY ON EMPLOYMENT ZONES

Overview

The following report presents the findings and recommendations from a study for Eurostat (the Statistical Office of the European Communities). Full details of the analyses undertaken as part of the research have been supplied to Eurostat on magnetic media. Each of the report's six sections addresses one important question:

1. Why are Employment Zones needed?
2. What are the requirements of Employment Zone definitions?
3. Can these requirements all be satisfied?
4. Which member countries have defined Employment Zones before?
5. Can one country's definitions be applied to other countries?
6. What are the recommended definitions for Employment Zones?

This report has been written by the NE.RRL project team* who acknowledge the inputs of Fabio Sforzi (IRPET, Florence) and the Eurostat officers Hubert Charlier, Benoit Paul and Vital Schreurs.

*Mike Coombes (Research Coordinator) Stan Openshaw (Technical Director) Colin Wymer, Martin Charlton, Simon Raybould.

Report from SOEC project number 0781021

1. The need for Employment Zones

The collection and analysis of local statistics allows the trends and conditions of each local area to be compared. Statisticians devote considerable time and energy to ensure that statistics are collected in a consistent way so that the data for different areas is genuinely comparable. In the European context, this has led to new datasets, such as the Labour Force Surveys, which allow unemployment levels and trends to be compared between (parts of) different countries. However, the need for local or subregional statistics raises the possibility that the very areas used for reporting the statistics may themselves undermine this comparability.

The problem is best explained by some examples. There is likely to be considerable interest in comparing London and Paris. In the case of London, the existing administrative boundaries offer two alternative definitions: one is the medieval City (with a population of around 3000), and the other is Greater London (virtually the whole built-up area, with a population of over 6000000). There are also two main administrative definitions of Paris: the Department is roughly half-way between the two London alternatives, the Region extends far beyond the built-up area. In short, no common approach exists between the two systems - and it is no easier to find an equivalent definition between either city and the Community's one other conurbation of a similar size, the Rhine-Ruhr area of Germany.

The reason why this difficulty is important follows from the purpose of Eurostat's analysis of local trends. One of the major policy tasks is the definition of areas eligible for "Objective Two" assistance. In practice, this identifies areas with particularly problematic economic structures. It is well known that different industries often cluster together within a wider area. As a result, to analyse the structure of the City of London alone would reveal an extraordinary predominance of the financial sector. In contrast, an analysis of Greater London would balance this picture by including large numbers of jobs in government and other service sectors. Thus the choice of area to analyse would greatly affect the results

obtained from the same dataset - a fact that would be all the more evident in areas such as Rhine-Ruhr where adjacent towns have different specialisms.

Another aspect of the problem is illustrated by the recent tendency for inner urban areas to experience decline while the same cities' suburbs and surroundings are growing. As a result, London and Paris could be experiencing very similar patterns of decentralisation, and yet show quite different statistical profiles: the Department of Paris might highlight the decline in the inner urban areas (perhaps with associated problems such a high unemployment), while Greater London's inclusion of more affluent suburban areas would provide a more balanced picture.

The implication of this problem is that the unquestioning use of administrative areas for statistical analysis is likely to prevent the results being genuinely comparable. Even if some member countries have administrative areas with boundaries that have recently been revised to make them geographically well-defined and consistent, they cannot be meaningfully compared to the areas of other countries. Consequently, the need for Eurostat to provide comparable analysis of all local areas for policy purposes inevitably leads to the need for a specially defined set of areas. This report proceeds to identify the appropriate basis for defining these areas, which are to be termed Employment Zones.

2. Definition of Employment Zones

This section identifies the key principles which need to be applied to ensure that each Employment Zones (EZ) is sufficiently comparable with each other EZ for their use in a policy context. The first principle is that the EZs are statistical areas - in cases where administrative areas do emerge as the best available EZs, then their suitability needs to be demonstrated statistically. The relevant statistical tests will be derived from the other principles of definitions, but the primary principle is that such tests are necessary to demonstrate comparability.

It is likely that many potential EZs will meet the statistical criteria without any difficulty: in these cases there will be 'room for manoeuvre' which will allow the detail of the EZ boundaries to be drawn in ways that more closely satisfy other considerations. This 'fine tuning' might take the form of manual adjustments to the outcome of a computerised procedure which defines the initial set of EZs that all meet the statistical criteria. The subsequent adjustment step would also offer an opportunity for local consultation and allow the definitions to take account of new developments since the data was collected for the main computer analysis. However any adjustments need to be limited by checking that the final EZs still meet the statistical criteria that were set to ensure that the areas meet the primary principle.

The second principle is that the concept underlying EZs is the local Labour Market Area (LMA). The fundamental feature of LMA boundaries is that they lie along 'traffic watersheds' and as such represent the dividing line between areas within which most people both live and work. Thus the definition of EZs needs to flow out of analysing journey to work data, in order to find boundaries across which relatively few people travel between home and workplace. The policy relevance of this principle can be seen as 'targetting' - the residents of a local area which is assisted to broaden its range of employment opportunities will not benefit much if journey-to-work flows are so diffused that many of the new jobs are taken by other areas' residents.

The third principle is partition: the apparently self-evident requirement that each part of all member countries should be in one and only one EZ. It is worth pointing out here that there is no clear 'natural' method for defining IMAs, so that there could be many alternative approaches to identifying EZs. However, this principle rules out those approaches which do not allocate to EZs the whole of the territory they analyse. It also rules out any system of overlapping areas - which includes any set of EZs that is dependent upon a 'two tier' structure of boundaries (additional 'tiers' may be considered separately, but a single tier of EZs must provide the full set of comparable areas for statistical and policy purposes).

As summarised in Table 1, the first three principles can be considered to be more fundamental than the others. The fourth principle is that each EZ should form a single continuous territory: that the internal contiguity of each EZ is ensured. Of course, the complicating factor of islands can make this principle difficult to operationalise routinely. It is also important that the requirement only applies to the final set of EZs - indeed, to apply a contiguity constraint throughout all the stages of defining EZs often distorts the final set of EZs by placing too much emphasis upon the detail of the local zone boundaries (which will often be arbitrary or anachronistic). The fact that relatively few people travel very long distances to work means that, for any area, most links are with adjacent areas. Usually, then, it is possible to deal with any non-contiguous parts of otherwise well-defined IMAs by minor manual adjustment at the end of a computerised procedure.

The fifth principle is that the EZs should maximise their autonomy in terms of internalising the flows between home and work. This self containment is the main statistical expression of the objective to define IMAs, in that it also minimises flows across the boundaries. However, the notion of self containment develops this general objective into twin criteria: that each EZ not only provides local jobs for most of its resident workers, but also provides local workers for most jobs in the area's workplaces. The statistical analysis leading to EZ definitions should ensure that every EZ meets minimum levels of self-containment. It is also possible to compare

Table 1 Principles for a common system of Employment Zone definitions

1. Statistical Areas:	every Employment Zone (EZ) has to be defensible as a consistently defined unit, so that the set of EZs are comparable for statistical and policy purposes
2. Concept:	the Zones should represent Labour Market Areas (LMAs) that were identified by appropriate criteria that are applied objectively in each part of the Community
3. Partition:	each area of the Community should be in one, and only one, Employment Zone (so that there is a single 'tier' of Zones that thereby covers the whole Community)

****

THE COMBINATION OF PRINCIPLES 1 TO 3 SETS THE OBJECTIVE OF THE EMPLOYMENT ZONES

****

4. Contiguity:	each Zone should be a single coherent territory (but this constraint may be imposed as a last step to include isles and to minimise the effect of the Local Zone boundaries)
5. Autonomy:	every Zone should meet a minimum self-containment level (so that most of the Zone's workers live in that Zone, and most of the Zone's employed residents work locally)
6. Homogeneity:	the comparability of Zones will be improved by making sure that they are of a similar size (in particular, a minimum size level should prohibit very small EZs)
7. Coherence:	the Zones should not often have elongated shapes, nor have boundaries that cannot be recognised as a reflection of local topography
8. Adherence:	the Zones will fit NUTS boundaries if possible, with greater emphasis on the top tier boundaries, but even national borders may need to be crossed
9. Flexibility:	the method will need to cope with such varying (eg. Paris, the Shetland Isles, Calabria) that it must not be based on just one model of LMAs

****

WHERE PRINCIPLES 4 TO 9 CONFLICT, PRIORITY IS GIVEN TO THAT HIGHER IN THIS LIST

****

one approach to EZ definition with another in terms of the overall self-containment of flows within each set of possible EZ boundaries.

The sixth principle is homogeneity - in that the comparability of EZs will be impaired if the areas are of widely divergent population size. The most straightforward expression of this principle is that EZs should not fall below a minimum size. The smallest areas are the most vulnerable to extreme and volatile statistical profiles which are thereby unsuitable for policy purposes. It is also preferable if EZs are not larger than is necessary to represent the pattern of IMAs: larger EZs would embrace several IMAs which may have quite different levels of need for policy assistance.

The seventh principle is coherence - EZ boundaries should not be unnecessarily complex. In general, the clustering of work journeys (around urban areas in particular) provides the foundation for basically compact IMAs. Even with the recent trend towards longer journeys to work, the majority of flows are quite localised and spatially focussed. As a result, EZ boundaries should usually conform to broad expectations based on local topography, route networks and settlement patterns. EZ boundaries that strongly diverge in ways that appear erratic or highly irregular may result from a method of definition that places too much stress on a minority of unusual flows (which may simply be from unreliable data).

The eighth principle is that adherence to standard administrative area boundaries is advantageous. Notwithstanding the preceding discussions, if there are two alternative possible sets of EZ boundaries which are otherwise equally acceptable, then the one that more closely matches administrative boundaries is to be preferred. The policy basis for this principle is self evident, as is the further implication that 'higher order' boundaries (eg. of Level I regions in the NUTS hierarchy of administrative areas) are the more important to be 'matched' by the EZs. However, even member state boundaries are not of over-riding importance, because it is known that flows across these frontiers are already substantial in some places, and are growing in most areas. Cross-border EZs are likely to remain unrecognised, however, because journey-to-work

data is generally made available only for a single member state (or even a region within it), so there is little consistent data on cross border flows. At the other end of the spectrum, the one form of area whose boundaries will have to be respected is the 'local zones' which, with their far smaller average size than Level III areas, have been adopted by Eurostat as the appropriate scale of analysis for defining EZs.

The ninth and final principle is that flexibility will be essential for a common method of defining EZs. The policy context into which the EZs will fit requires that the EZs be accepted as an adequate representation of LLMAs by local and national experts. One aspect of flexibility might therefore be that the definitional process includes an opportunity for evaluation and adjustment, although this must be constrained so that no EZ consequently fails to meet the statistical criteria which ensure a basic level of comparability between areas. The other aspect of this challenge follows from the dramatic variation between LLMA patterns in such strongly contrasting regions as southern Italy, northern England and the Danish islands. For example, some areas have commuting flows strongly focussed on certain foci (eg cities), others do not. The need for statistical comparability denies the possibility of varying the statistical criteria between one region or country and another. The flexibility required here has to be inherent in the method of definition - it must be able to distinguish the major pattern in a set of local commuting flows, whatever that pattern may be. In practice, the procedure for defining EZs will need to be very highly generalised if it is to generate EZ boundaries that pass critical local and national scrutiny in all the diverse circumstances across member countries.

3. Putting the Principles into Practice

The nine principles that have been outlined above were summarised by Table 1. The first three - defining a set of statistically valid labour market areas in a single complete tier - can be taken as absolute requirements which are logically compatible with each other. Together they define the objective for the EZ definitions. This objective should be achievable unless no procedure for defining statistically valid EZs can be developed, or labour market behaviour is so complex that it cannot be adequately represented in a single tier of EZs. This report has already asserted that the latter possibility is not the case in Europe, the remainder of the report will establish the outlines of a solution to the former challenge.

Whereas principle nine essentially summarises the difficulty faced in seeking a common approach, principles four to eight provide the crux of the methodological challenge of EZ definitions. Individually, each of these principles is an expression of one aspect of the first three principles that set the objectives for the definitions. For example, the preference for similarly sized EZs follows from their use in policy analysis, which requires that the areas should have as comparable a statistical basis as possible. However, the uneven settlement pattern across member countries ensures that meeting the second principle - identification of labour market areas - guarantees that the EZs will in practice have substantially varying population sizes. As a result, the sixth principle only requires that this variation should be minimised - not altogether prevented. The critical point here is that the statistical requirement to align EZs with LMAs (principle five) is a higher priority than that (principle six) of minimising EZ population size variation. This key point is expressed in Table 1, where the principles were listed in descending order of priority.

The high priority afforded to EZ contiguity (principle four) need not lead to it over-riding all other considerations. The previous section of this report stressed that contiguity tends to emerge 'naturally' as a result of the clustering of work journeys in most localities. Consequently, contiguity can be ensured as part of a final stage of the definitional procedure - including the adjudication of complex issues such as the

treatment of islands. To restrict the analysis by enforcing contiguity from the first stages of the definitional procedure places too much emphasis on the vagaries of the local zones' boundaries. Not imposing contiguity in the early steps of the analysis allows the definitions to recognise that an EZ may have several focal points that are not contiguous (eg 'twin cities' or the dispersed foci of coal mining areas). This approach should not lead to major centres grouping together (as a result of long-distance commuting, eg between Lyon and Paris), provided that the definitions use appropriate methods: even if a zone does have a large number of long-distance commuters, the majority of workers will still work locally.

The fifth principle is the one which provides the critical statistical 'tests' for the EZ definitions: self-containment. The previous section of the report stressed that these tests should both provide a guarantee that no EZ fails a minimum level of validity and provide a basis for evaluating one set of such EZs against another. Thus the statistical procedure for defining EZs can be 'optimised' in cases where local zones could validly be allocated to more than one EZ (ie. they should be assigned to whichever leads to higher overall levels of self-containment). In other words, the minimum level of self containment for EZs sets a limit to the 'room for manoeuvre' within which the procedure can also use self-containment levels to optimise the EZ boundaries.

The fact that there can remain alternative sets of EZ boundaries, each of which meet the minimum self-containment levels, means that the lower priority principles (six to eight) remain relevant. There are two ways in which they can be brought into the EZ definitional procedure. They can be added as a final stage, perhaps as part of a consultation process with local and national experts. However, it would be unreasonable to expect an unstructured procedure to consistently adjudicate on the conflicting priority of several different principles in each case. The other alternative is to express (some of) these principles as statistical tests, in addition to the self-containment test. The definitional procedure can then combine these various tests into a single analysis, within which the

relative 'weight' given to each test reflects the relative priority of that principle (as shown in Table 1).

An example can illustrate these technical options best. Principle eight is that adherence to administrative boundaries is an advantage. The low ranking of this principle emphasises that it should not take priority over any of the statistical principles for the EZ definitions (Table 1). However, there will be numerous areas where there are several alternatives for the detailed alignment of boundaries and where more than one alternative set of boundaries satisfies the statistical criteria. A sophisticated analysis will be able to identify these options and to automatically select that option which best 'fits' with major administrative boundaries in the area.

A less sophisticated approach will simply identify the set of boundaries that is 'best' on the statistical criteria, but will then provide a testing procedure so that adjustments to the boundaries can be assessed interactively. Thus re-aligning the EZs so that they match administrative boundaries would be tested to see if this option is within the 'room for manoeuvre' that is limited by the statistical criteria. If the adjustment would cause any of the EZs to fail these criteria then it would not be accepted. However, a change could be accepted if the areas' critical values were lowered (eg by becoming less self-contained), but were still above the critical level: in such cases the advantage of alignment with administrative boundaries could be considered to outweigh the statistical disadvantage of the change.

The consultation stage of the British TIWA definitions was undertaken with this type of consultation and adjustment stage following the computerised analysis. One other lesson learnt then was that imposing 'higher order' boundaries at the outset leads to less satisfactory IMA definitions. In particular, the data for Scotland and N. Ireland were both entirely separate, so that it was impossible to recognise any flows across the English-Scottish border (which cuts through some genuine IMAs). Imposing other boundaries, such as that between England and Wales, only increases this problem. The long-term objective must be for a single dataset

covering all member countries. For the present, no existing dataset should be sub-set for the analysis, because this sub-setting inevitably distorts the EZ boundaries that result. If the EZs are to be aligned with administrative boundaries, then this should be tested in one of the ways set out above - it should not be imposed by sub-setting the data.

In the following sections there will be some discussion of methods that demonstrate the feasibility of combining several criteria into a single procedure. Even principles such as the need to avoid irregular boundaries can be represented by statistical tests (in this case, maximising the compactness of the EZs). However, it is likely that one or more of the principles will remain to be implemented manually - as may the application of the contiguity constraint to island areas. These refinements can be combined with wider consultations to achieve acceptability of the resulting EZ boundaries. However, the possible boundary changes considered in this final stage should be constrained to within the 'room for manoeuvre' identified by the statistical criteria. In summary, some contradictory criteria can only be resolved through value judgements, whereas others can be expressed within statistical tests that should then be used to ensure that all defined EZs meet minimum levels of comparability.

4. Existing National Approaches to IMA Definitions

Governments in several member nations have used IMAs for reporting official statistics and implementing certain policies. Most unusually for official boundaries, IMA definitions in a number of several countries have depended heavily upon research by academics. Of course, any such definitions depend upon the availability of data on journey-to-work flows at the very local level. To date this constraint has prevented the possibility of IMA definitions in Ireland, Portugal, Greece and most of Spain. There have also not been official IMA definitions in Denmark, Belgium and Luxembourg, although the relevant data there has been used to inform other procedures such as the revision of local government boundaries. Thus the countries where there are official IMA definitions are France, Germany, Italy, Netherlands and the United Kingdom. Each country has developed its own approach:

Dutch : administrative analysis of commuting data defines the COROP
British: commuting data is used to identify Travel-to-Work Areas (TTWAs)
French : commuting data is used in the Mirabelle method
German : two methods are combined to define Regional Labour Markets
Italian: a variant of the British method has been devised (NIRA)

The data that was needed depended upon the detailed requirements of each method. The administrative component in the Dutch method was one reason for this approach being too difficult to apply directly to other countries. However, the other national methods are sufficiently similar for comparisons to be made. It is not necessary here to describe in detail the differences between the existing national methods. Table 2 summarises the major issues involved in defining IMAs, focussing on eight key questions:

* Does the method start by defining 'cores' (usually, the main urban centres) around which to build the IMAs, and if so, how are they identified?

* When linking a zone to a core or a set of zones, on what basis is the most appropriate linkage chosen?

* Is there a contiguity constraint (ie. a restriction on the analysis so that it only considers zones that are neighbours) when linking zones into IMAs?

* Given that the sequence of analysis partly influences the results, how does the method select the order in which to 'build' the IMAs?

* Given that any sequence of analysis will be suboptimal in some areas, does the method consider changing some linkages at a later stage or is it rigid?

* A key feature of IMAs is self-containment (ie. the proportion of the area's employed residents who work locally, and the proportion of jobs at local workplaces filled by local residents) - does the method set minimum levels?

* Does the method set a minimum population size level for the IMAs?

* What manual 'fine tuning' is applied to the method's computerised results?

As shown in Table 2, each method has some particularly distinctive features in its approach to these questions:

French - an elegantly simple approach in which the linkage procedure is constrained by contiguity and the linkage is hierarchical.

British - a complex method with the extra refinement of a step to 'optimise' the IMA boundaries, but with the contiguity constraint applied manually at the end.

Italian - a close derivative from the British method, but with the major additional step or re-running the whole computerised procedure so that the IMAs are part of a two tier system of regions.

German - a totally distinct approach which initially involves two wholly separate analysis: the results from the two parts are then combined to impose a limit on intra-IMA travel time.

Table 2 Comparison of existing national methods

Critical features of each method	French method "Mirabelle"	British method "TTWAs"	Italian method "NIRA"	German method part A	German method part B
Are 'cores' selected as a starting point?	not done	EITHER foci of net in-commuting OR highly self-contained	foci of net incommuting	separate foci of net in-commuting	foci of incomm-uting
How is the best link chosen for each zone?	measure of dependency in terms of commuting (asymmet-rical)	measure of inter-dependency in the flow data (symm-etrical)	(as British method)	Reilly's Law, but limited by travel time maximum	size of flows (asymm-etrical)
Are the links rest-ricted by contiguity?	explicit restriction on links considered	computer analysis is not at all restricted	(as British method)	implicit to use of road network	not re-stricted
How is the sequence of analysis decided?	descending order of the values of all the feasible links	complex multi-step procedure, derived by experim-entation	(as British method)	descen-ding order of cores' strength as foci	not explicit
Is the aggregation procedure sensitive?	rigidly hi-erarchical	hierarch-ical at first but later steps optimise the regions	(as British method)	rigidly sequen-tial	rigidly hierar-chical
Is there a self-containment minimum?	resident self-containment only	resident AND work-place self-containment ....but....	workplace self-containment only ....but....	not explicit	not explicit
Is there a population minimum?	a variable constraint	population size is in a combined 'target' with self-containment	(as British method)	no	yes
What 'post-processing' is applied at the end?	the results are subject to regional evaluation	the areas are subject to detailed consult-ation (eg. to impose a contiguity constraint)	the areas defined by computer are subject to a second analysis to create an Upper Tier	the sets of areas resulting from the separate part A and B analyses are combined into one set of boundaries	

5. Experimental comparative analysis

A comparison of each method's applicability to the other member countries was the first step towards finding a common Community approach to defining Employment Zones. The test of the German method has been frustrated by the lack of travel time data in other countries. The two tier feature of the Italian method is not appropriate to the definition of EZs: the method is otherwise close to the British method so that a single set of experiments can illustrate the approach here.

5.1 Available data

The British method only required commuting data, contiguity data was also required for the French method. Table 3 specifies the data available in each member country. An important feature of any commuting dataset is the size of the zones that make up the matrix. In general, the larger the zones' population, then the higher the share of flows that will be contained within individual zones. Also, a matrix of large zones will tend to include flows which are larger and more 'regular' (eg tending to show that most non-internal flows are to nearby/contiguous areas). It is therefore crucial to recognise, when comparing the results in the different countries, that for example the data zones in Denmark are on average over 10 times larger than the local zones in France (Table 4).

Two further points are relevant here. The first is that the Italian data has been subset, by deleting all flows involving a single person. This procedure (to reduce the dataset's size) should not be necessary in future, given the powerful hardware used by NE.RRL for this project. In fact, experiments with another NE.RRL dataset, which includes all the single person flows, produced very similar results. This finding could be seen as an encouraging, if indirect, sensitivity test.

The second noteworthy feature is the Belgian data's inclusion of flows to nonBelgian destinations. The inclusion of this outflow data is very unusual, as shown by the fact that even flows between England and Scotland are uncoded in the British data. The inclusion of international flows is

Table 3 Available information for this research

Country (coverage)	Local Zone commuting data (special features)	Boundary data	Population & Employment Census data
Belgium (all)	Yes (also flows to non-Belgian destinations)	Digitised in NE.RRL	No
Denmark (all)	Yes (Danish local zones are large)	Digitised in NE.RRL	No
Germany (Western Lander)	No (data supplied in wrong format to Eurostat)	No	No
Spain (Catalonia)	No (but data has been supplied to Eurostat)	No	No
France (mainland)	Yes (French local zones are very small)	Contiguities only	No
Greece (all)	No (no commuting data collected)	No	No
Ireland (all)	No (data may not correspond to local zones)	No	No
Italy (all)	Yes (all flows of a single person deleted)	Contiguities only	Yes
Luxembourg (all)	No (cross-border flows a major issue)	No	No
Netherlands (all)	No (data considered unsuitable by NSI)	No	No
Portugal (all)	No (no commuting data collected)	No	No
United Kingdom (England & Wales)	Yes (data is a 10% sample)	Yes	Yes, plus other data

Table 4 Details of commuting data for the comparative research

Dataset	Area (,000 km2)	Pop.n (million)	Exceptional features of the dataset	Local Zones [ave. pop.n (,000)]	Non-zero flows ave. no. per Zone	Non-zero flows % of all that = 1
Belgium	31	10	flows to (but not from) other countries included	589 Communes [17]	89	38%
Denmark	43	5	none	276 Kommuner [18]	91	37%
England & Wales	150	50	10% sample [different to data for TTWAs]	9689 Wards [5]	47	50%
France	549	56	none	36082 Communes [2]	31	53%
Italy	301	57	excludes all flows that were of a single person *	8086 Communi [7]	38	* 0% [44% in original dataset]

very important given Belgium's geography: it would be almost as critical for the Netherlands, and far more so still for Luxembourg. However, the real need is for data on flows in both directions, but inflows are not coded for Belgium (or any country). This objective will only be met when every National Statistical Institute (NSI) codes fully the destinations of all their residents.

The potential importance of cross-border data, given 1992 developments, is great. Even the 1981 Belgian data shows that well over 10% of all Communes had over 10% of their residents commuting to jobs outside Belgium. This leaves a genuine dilemma because the set of purely internal LMAs may appear to be a less reasonable portrayal of Belgium's geography, whereas the results from the including the outflows in the dataset is less comparable to that of the other countries (and still omits the 'inflows' that are likely to be equally significant to the overall pattern). Because the research here involves comparing results from different methods, the same dataset must be used for all methods, so the necessary 'standard' datasets are those that have been distributed for analysis by the different methods. The point to stress is that it would be possible to remove some of the special features of individual countries' datasets documented in Table 3 and 4, if the purpose here was to assess more comparably the value of any one method when applied in each country.

5.2 Initial results

Table 5 shows the results from applying the basic French and British methods to each of the five countries' datasets. The results shown first from the British method in England and Wales are not identical to the official TTWAs because the latter included the effects of consultation on the results of the computerised analysis, notably to remove non-contiguities from the final boundaries. In fact, non-contiguities in the results are only numerically significant in France, where the small size of the local data zones exacerbates the problem. The second column in Table 5 shows the results from another run of the British method, but using a higher 'target' for the statistical criteria in order to generate fewer LMAs.

Table 5 Results of the comparative research

(self-containment %)	British method		French method	
	as TTWAs (70 -> 75%)	high target (70 -> 95%)	"70% active"	"40% migrants"
Belgium	24	19	22	20
Denmark	40	31	41 *	34
Eng. & W.	247	180		
France	859	383		
Italy	496	282	269	235

* in this case, the setting was 80% self containment

The other columns of Table 5 show two sets of results from the French method. The "70% self-containment" analysis appears to have quite a similar 'target' to the original British analysis, yet generates far fewer IMAs in Italy. The last column shown uses a somewhat different form of self-containment" test, but identifies numerically similar results to the higher 'target' version of the British method. These two runs form the basis for the further comparisons made later in this report. It is an obvious disadvantage here that no results have yet been made available from applying the French method to England & Wales or, indeed, France.

It is clear from Table 5 that the size of the local zones is not the only factor that determines the number of IMAs identified in any particular analysis. The evidence of Table 5 is somewhat over-simplified however, because the French method is not designed to be implemented universally in any application. The approach is usually to generate several alternatives, then to select the one that most closely accords with local knowledge. Such an approach is strictly inconsistent with the emphasis here upon comparability of EZ definitions across all member nations. Thus the following discussion will assume that a single set of criteria is applied consistently.

On this basis, Table 5 suggest that both methods identify numbers of IMAs in each country that broadly reflect its level of urbanisation and internal cohesion. Thus there are few IMAs covering the highly integrated Belgian territory, while the large rural tracts of Italy and France are reflected by large numbers of separate IMAs - particularly with the lower 'target' criteria settings.

5.3 Sensitivity analysis

In the absence of the French results for two countries, this sub-section only considers the British method. Table 6 expands on the first column in Table 5 by showing the proportion of each country's TTWAs that have certain characteristics. The TTWA population values show that Belgian's high level of integration has not led to just one or two large TTWAs, but that over

Table 6 Basic results from the TIWA method

	No. of TIWAs	% TIWAs with resident self-containment		%TIWAs >40,000 workers
		<75%	>95%	
Belgium	24	38	0	71
Denmark	40	5	8	40
England & Wales	250	20	3	46
France	898	6	7	10
Italy	509	4	16	13

two thirds of the TTWAs have over 40,000 jobs (a size that identifies areas likely to have resident populations of over 100,000). Unsurprisingly, the very large number of TTWAs in the two Mediterranean countries include only a small proportion that have large populations. The TTWA self-containment data in Table 6 gives the rather surprising result that although France has by far the largest number of TTWAs, it is Italy which has the highest proportion of TTWAs that are very highly self-contained. The exclusion of single person flows from the Italian dataset can only partly explain this result. The very high number of TTWAs identified in France will be slightly due to the very small data zones there (because small zones allow the method to identify a large number of TTWAs that only just satisfy the basic population and self-containment requirements).

One form of sensitivity analysis is to change the settings on the 'target' criteria and then observe the effect on the results. It is most appropriate to raise the 'target' settings, because the results in Table 6 include an unwieldy number of TTWAs in some countries. The higher 'target' values, which produced the second column of results in Table 5, ensure that any very small LMAs which continue to be recognised separately must be very highly self-contained. This adjustment is not claimed to provide the ideal set of areas: it is simply a 'move in the right direction' which shows the sensitivity of the method's results in different countries.

Table 7 shows that this change makes a substantial impact on the overall results. The reduction in number of LMAs produced varies between just over a fifth (Belgium and Denmark) to over half (France). It is encouraging that the largest reductions are on those countries which had the largest - and probably least justifiable - number of TTWAs in Table 6.

The other information in Table 7 provides a classification into five categories of the LMAs produced. The first such column includes only LMAs that are 100% self-contained: these have no data on flows to or from other parts of their country and in all cases have populations too small to be of interest. The method is therefore unable to allocate them with other parts of the country; they are 'failed' LMAs that will appear separately in any analysis (including all the other Tables here).

Table 7 Results from the sensitivity analysis

	No of LMAs (as % of (as % basic results)	% Areas with no. employees [E] and resident self-containment [R]:				
		E <2500	E ≥2500 <4000		E ≥4000	
		R = 100%	R <90%	≥90%	R <90%	R ≥90%
Belgium	19 (79)	0	5	0	84	11
Denmark	31 (78)	0	32	13	35	19
England & Wales	183 (73)	2	22	7	55	15
France	422 (47)	9	32	21	8	30
Italy	291 (57)	3	24	37	12	24

The remaining four columns of Table 7 cross-classify the other LMAs into higher/lower population and self-containment groups. Belgium is clearly unusual, having so many LMAs in the category of high population and low self-containment. This column (the last but one) provides the strongest contrast between countries, with England and Wales most similar to Belgium, while France and Italy have very few LMAs of this kind (which are mainly found in suburban or industrial districts). Table 7 also confirms the earlier evidence that Italy has a high proportion of LMAs with small populations but high self-containment.

Another assessment of stability can be derived from a comparison of the boundaries of the LMAs produced by these two different versions of the British method. In detail, the stability of the boundaries is not particularly high: this is because the British method is non-hierarchical and seeks to 'optimise' the boundaries for that specific set of LMAs. For example, the boundaries between three similarly sized centres (A, B, C, set in a triangle) will probably differ depending on whether the three are each in a different LMA or not. If they are, then the boundaries will probably lie close to the midpoint between each centre. However, if B and C are grouped into a single LMA, then their common boundary should probably lie nearer to A; this is because the intervening areas will have flows with both B and C which, in combination, tend to outweigh the flow with A alone (except in the zones very close to A itself). A strictly hierarchical method will simply remove the boundary that had previously separated B and C - it would fail to re-assess the optimal location for their (now) common boundary with A.

This detailed 'instability' of boundaries becomes more obvious where there are complex urban systems, and where there the dataset is for such small zones that the 'fine tuning' of the boundaries can be observed. The opposite case is provided by Denmark, where the zones are large and the urban system fairly clear. In this case, all but two of the boundary changes, between the set of results shown in Tables 6 and 7, is a straightforward removal of a previous boundary. In other words, seven of the nine LMAs that are 'lost' (with the raising of the settings for the criteria in the method) are wholly merged with adjacent LMAs. These

results show that where there is a clear pattern in the flow data - as in Denmark - the British method succeeds in portraying that stability. Consequently, the 'instability' of the British method's results elsewhere is likely to be an accurate reflection of the complexity of local commuting patterns there.

5.4 Types of area

There are a number of geographical circumstances that pose problems for the definition of labour market areas. These types of region are now discussed in turn, with an example taken to discuss the validity of the British and French methods' results in each case where appropriate.

Metropolitan regions can prove problematic, particularly where they are surrounded by no very substantial urban centres. Copenhagen provides a clear example: here the British method defines a large LMA centred on the capital, but succeeds in limiting it to less than the full extent of Sjaelland. The set of results with the reduced number of LMAs (Table 7) has succeeded in removing nine smaller TTWAs in more rural parts of Denmark without significantly increasing Copenhagen's size. In contrast, the French method creates on LMA based in Copenhagen that extends to the limits of Sjaelland.

Conurbations, with many adjacent similarly sized centres, are found in many industrial parts of England and Wales in particular. The British method has been devised to repeatedly re-assess the most appropriate set of boundaries from among this complex pattern. Thus the number of final areas may be changed with the criteria settings, but the new set of boundaries are re-optimised accordingly. A hierarchical approach such as that in the French method becomes increasingly sub-optimal as the process continues. The pattern of boundaries around Manchester, and the 'satellite' towns that form a ring around it from Stockport to Bolton, provide a good example of the TTWA method's success in these circumstances.

Large rural regions provide a parallel problem to the previous type, except with far fewer flows in general. The LMAs defined in Brittany by the

British method have been acknowledged by the French NSI to succeed in recognising most local centres and their hinterlands. Again, the benefits of the non-hierarchical approach are important to ensure that appropriate boundaries are still being defined after a long sequence of analysis steps, a situation that often arises in rural regions where very many zones have to be grouped in order to reach a minimum size criterion.

Islands can be difficult, especially for methods dependent upon contiguity information. The results in the physically fragmented parts of Denmark show that that the British method is able to identify coherent patterns in the commuting flow data (which shows linkages between zones, whether or not the links are between zones that are separated by water). The French method tends to be restricted to identifying islands as separate IMAs.

Border areas are a peculiar difficulty, as mentioned in several sections of this report. There is really no entirely satisfactory solution possible in the absence of full cross-border flow data. The evidence in Belgium does, however, provide some encouragement in the British method's ability to distinguish the significance of smaller centres (eg La Louviere) even where these are over-shadowed by the influence of several larger centres (Mons, Charleroi) and also lie close to an international border (in this case, of France) so that their local flow data will be incomplete. Some of the French method's runs recognise a La Louviere IMA, but the "40% migrants" analysis (which is being taken here as a 'standard' run) groups it with Brussels some distance away.

A final interesting case study is provided by the industrial towns of Pistoia, Prato and Empoli which all lie within 25 miles to the west to Firenze. They have distinct industrial specialisms, and are widely discussed as maintaining quite separate local economies, despite their economic growth in recent years. In the NIRA 'lower tier' each is the centre of a locality, but at the 'upper tier' Firenze absorbs them - except for Empoli which groups with the rather distant Pisa. The 'standard' French run ("40% migrants") groups together Pistoia and Prato but not the other two (some other variants on the French method maintain all four separately). As for the British method, both the runs that are being

considered here identify the four individually as the centres of separate IMAs - the result which is probably the most justifiable on the basis of the wider evidence in the literature on Tuscany's local economies.

6. Conclusions

The research findings are best summarised by reference to the principles which have been outlined for a common system of EZ definitions:

1. EZs should have the statistical basis needed for policy application -

   *the common method should ensure that its results will conform to the statistical requirements imposed by the areas' use for policy; it is also recommended that the British approach of identifying any 'room for manoeuvre' is adopted so that some manual adjustments can be carried out (to match the results closer to local knowledge) while keeping within the statistical criteria that have been set at the outset.

2. The method of definition should identify labour market areas -

   *all the current national methods have been evolved within programmes of research on local labour market areas.

3. Each local zone to be in 1 and only 1 EZ within a single tier -

   *all the basic methods guarantee this outcome (subject to problems within the data, eg zones without any flows in or out), but the Italian approach is designed with the aim of identifying two 'tiers' of regions.

4. Each EZ to comprise a single contiguous territory -

   *the recommended method will generate LMAs that are more optimal statistically if it does not have a contiguity constraint within its initial stages (the original TTWA method depended on the later consultation stage to impose contiguity manually, but the British software has now been enhanced with a 'contiguiser' as the final step of its program).

5. Commuting flows should be as self-contained as possible within EZs -

*the recommended method will be non-hierarchical, because otherwise results become less optimal in terms of self-containment; the same criteria should of course be applied to all member countries to ensure comparability.

6. Population size range to be minimised -

*the critical parameters in the method should explicitly encourage the setting of a population minimum (but a maximum should not be imposed).

7. Boundaries should usually appear to be coherent and recognisable -

*the clustered nature of commuting patterns tends to reduce the likelihood of IMAs with very irregular boundaries; a small number of anomalies can be dealt with by minor manual adjustments (but if necessary the objective of 'compactness' could be built into the critical parameters directly).

8. Adherence to the higher level of NUTS boundaries is preferable -

*where the original results are close to achieving this adherence, proposed changes could be checked to ensure that the resulting EZs still meet the statistical criteria; the relatively low priority of this objective reflects the preference for the original analysis to be constrained by as few boundaries as possible (ideally even spanning several member countries).

9. Method needs to be highly flexible -

*the earlier discussion of results in sample regions showed that the British method has the potential, after some experimentation with its 'target' values, to recognise the important patterns within flow data across very different types of geographical circumstances (there is not sufficient evidence yet on the results from other methods across a range of countries).

These nine requirements are all serious constraints on the development of a common approach to EZ definition. Experimental analysis with the British method has shown that it is possible to find adequate EZ boundaries in the countries for which data is available. The method has already been adapted so that it can analyse all 36082 French Communes simultaneously. Further evaluation is needed, most importantly to find the appropriate settings for the 'target' values (most notably, the population and self-containment crieteria) to bring the results closer to Eurostat's guidelines for the number of EZs needed in each country (Table 8). These experiments should prevent the need to adopt different definitional criteria in each country. The results could then be subject to national and local consultation, which should be constrained to maintain the statistical comparability of the EZs. This broad approach has been proven by experience with the British method, which also has methodological advantages in being non-hierarchical, and now also has an option to automatically impose a continguity constraint as a last stage. This last refinement has removed the final obstacle to offering this software as a new ERA (European Regionalisation Algorithm).

Table 8 Initial proposals by Eurostat for Employment Zones

COUNTRY	Eurostat proposal for number of EZs
Belgium	20
Denmark	20
Germany	200
Spain	250
France	300
Greece	50
Ireland	38
Italy	170
Luxembourg	3
Netherlands	20
Portugal	50
Britain	150