Socio-Economic Planning Sciences 47 (2013) 247–257

Contents lists available at SciVerse ScienceDirect

Socio-Economic Planning Sciences

journal homepage: www.elsevier.com/locate/seps

Optimal administrative geographies: An algorithmic approach

D. Datta [a], J.R. Figueira [b,c], A.M. Gourtani [d,e], A. Morton [d,*]

[a] Department of Mechanical Engineering, Tezpur University, Tezpur 784 028, India
[b] CEG-IST, Instituto Superior Técnico, Universidade Técnica de Lisboa, Lisbon, Portugal
[c] LORIA Laboratory, Nancy, France
[d] Management Science Group, Department of Management, London School of Economics and Political Science, London, UK
[e] Operational Research Group, School of Mathematics, University of Southampton, Southampton, UK

ARTICLE INFO

Article history:
Available online 11 April 2013

Keywords:
Primary Care Trusts
Multi-objective optimisation
Genetic algorithm
Healthcare geography
GP consortium

ABSTRACT

Centrally planned Beveridge healthcare systems typically rely heavily on local or regional "health authorities" as responsible organisations for the care of geographically defined populations. The frequency of reorganisations in the English NHS suggests that there is no compelling unitary definition of what constitutes a good healthcare geography. In this paper we propose a set of desirable objectives for an administrative healthcare geography, specifically: geographical compactness, co-extensiveness with current local authorities and size and population homogeneity, and we show how these might be operationally measured. Based on these objectives, we represent the problem of how to partition a territory into health authorities as a multi-objective optimisation problem. We use a state-of-the-art multi-objective genetic algorithm customised for the needs of our study to partition the territory of the East England into 14 Primary Care Trusts and 50 GP consortia and study the tradeoffs between objectives which this reveals.

1. Introduction

One of the more depressing features of health policy in publicly funded systems is the frequency with which reorganisations take place, redefining institutional roles, centralising what was previously decentralised (or *vice versa*), splitting or consolidating different delivery organisations, and so on [1]. These reorganisations have a substantial impact on the careers of healthcare workers, and on the quality of care experienced by patients, and so it seems desirable that they should be informed by the best possible evidence and science.

The current paper aims to provide a basis for more rational decision making around one aspect of reorganisation, namely the definition of geographically based entities which we call health authorities ("HAs"). We do not seek to prescribe whether Departments of Health undertake reorganisations. However, should a decision be made to reorganise, one might at least hope that the reorganisation will be made in a systematic and methodical manner. We present a multi-objective framework for making this sort of decision, and demonstrate a powerful multi-objective optimisation-based technology which can support such decisions.

Our analysis applies in particular to Beveridge systems, that is to say, National Health Service-type systems, rather than social insurance or health savings account systems. In such systems, administrative geography plays a critically important role. Typically, government funds are disbursed through geographically defined HAs, often according to a funding formula, which attempts to correct for differences in need between HA resident populations [2,3]. These entities are then responsible for providing care for their populations, either through hospitals and other acute facilities which they operate themselves or which they purchase from healthcare providers.

We will refer to a partitioning of a territory into HAs as a "healthcare geography". The question of how big these constituent units should be has been much debated in the literature [4–6], but not conclusively. Bojke, Gravelle and Wilkin's [5] view that "despite the importance of the issue of organisational size, the evidence available from published research is limited both in quantity and relevance" still seems to be true. However, the question – What constitutes a good healthcare geography? – seems to have received even less attention. In this paper, we tackle this more fundamental question directly. Having listed at the conceptual level some candidate attractive characteristics ("objectives"), we then turn our attention to the construction of metrics to operationalise these

* Corresponding author. Tel.: +44 020 7955 6537; fax: +44 020 7955 6885.
E-mail addresses: ddatta@tezu.ernet.in, datta_dilip@rediffmail.com (D. Datta), figueira@ist.utl.pt (J.R. Figueira), arash.gourtani@gmail.com (A.M. Gourtani), a.morton@lse.ac.uk (A. Morton).

0038-0121
http://dx.doi.org/10.1016/j.seps.2013.03.002

objectives, and formulate the question of how a territory should be partitioned up into HAs as a multi-objective problem. A multi-objective problem is a generalisation of an optimisation problem in which there are several conflicting objective rather than a single uncontested one. Rather than having optimal solutions, such a problem has (Pareto) efficient solutions: by definition, no efficient solution is better than another on all objectives. Then, for the empirical part of the paper, we turn our attention to revealing the tradeoffs imposed by an actual healthcare geography, in this case the geography of the East of England.

Because this problem is computationally intractable, we use a state-of-the-art multi-objective genetic algorithm customised for the needs of our study. Conceptually a multi-objective genetic algorithm works as follows. We have a very large set of feasible solutions, characterised in some suitable way and are interested in identifying efficient solutions. We proceed as illustrated in Fig. 1, which shows points located in a bi-objective space: we wish to minimise these objectives. First, we identify a set of starting solutions, shown as white circles. We perform certain operations on these solutions in order to identify a new set of better solutions (the grey circles). We iterate these operations and obtain the solutions represented by the black circles. If we iterate sufficiently, and our operations are defined appropriately, we will obtain a set of solutions which are efficient (or very close to being efficient) and which are widely distributed in the objective space.

Our paper is structured as follows. In Section 2 we discuss the current administrative healthcare geography of England, and some conflicting objectives by which it might be judged; in Section 3, we present the formulation of our multi-objective problem, our data sources, and solution method; in Section 4, we present results obtained from applying this model to a particular healthcare geography, namely that of the East of England; in Section 5, we conclude.

2. Background to the English system

In England, most healthcare is delivered by the government, funded out of general taxation. Health services are provided through the National Health Service (NHS): at time of writing, within the NHS there are two principal tiers of organisations with jurisdiction over geographically defined territories: Strategic Health Authorities (SHAs) at the top level and Primary Care Trusts (PCTs) at the second level. SHAs have a planning, coordination and monitoring role, and are effectively a local arm of the Department of Health; their structure almost exactly mirrors that of Government Office Regions. Primary Care Trusts or PCTs, of which there are 152, have responsibility for the delivery of healthcare to a geographically defined population. They directly control some 80% of the NHS budget and purchase health services from other NHS organisations, on behalf of their populations [7].

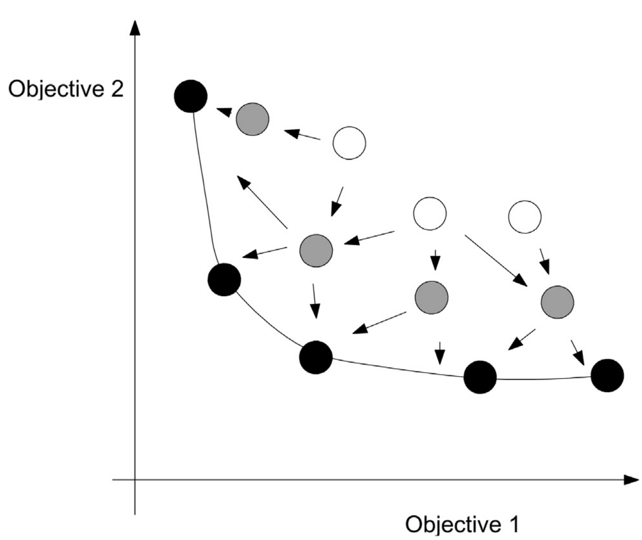


Fig. 1. Schematic illustration of a multi-objective genetic algorithm.

Recently, the incoming Conservative-Liberal coalition government has announced its intention to dissolve these structures [8,9]. In the future, PCTs and SHAs will no longer exist and funds will flow directly to consortia of General Practitioners (GPs) who will take over the commissioning of health services from Primary Care Trusts. The size of these GP consortia is not clear at the time of writing: however, it seems likely that the consortia will be smaller than PCTs, as the motivation for the change is to bring money closer to the individual patient. The critical questions of how money is to be allocated to GP consortia and how performance is to be managed under the new arrangements, have not been answered in any detailed way. This is not a trivial issue, as the heterogeneity of PCTs make both resource allocation and comparison of performance major analytic challenges. The statistical difficulties associated with devising performance management systems and resource allocation formulas for smaller and more diverse entities will represent a substantial challenge for analysts.

We suggest that one reason for the repeated reorganisations which have been a feature of life in the NHS in recent years might be that the question of what characterises a good system of HAs is intrinsically a multi-objective one. If this is the case, perennial dissatisfaction is to be expected – immediately after a reorganisation driven by some one particular objective, the failings of the new configuration with respect to all the other objectives will rapidly become apparent. Moreover, it is in the nature of policy processes in large bureaucratic organisations that even in the course of the same reorganisation, different objectives will be invoked by different actors to justify reorganisations in different parts of the system, resulting in an overall reconfiguration which cannot coherently be justified with respect to any set of objectives.

This paper deals with an aspect of the organisation of a healthcare system, namely the partitioning of a territory into HAs to form a “healthcare geography”. We propose that there are four broad characteristics of a healthcare geography which make it more or less attractive. We will refer to these as the “objectives”. They are:

O1. HAs should be geographically compact. As the HA is a single administrative unit and as health services are delivered at multiple points, staff or providers of commissioned services may be required to travel within the HA. Geographical compactness is attractive because it minimises travel distances. To make the point in a stylised manner, consider two HAs with area of 10,000 square miles, one of which, A, is constructed as a (compact) 100 mile square, the other of which, B, is constructed as a (non-compact) 10 $\times$ 1000 mile rectangle. The greatest point to point distance in HA A is $100 \times \sqrt{2}$ or 141 miles whereas the greatest distance in HA B is clearly in excess of 1000 miles.

O2. HAs’ boundaries should be roughly coextensive with other geographically defined entities with which they have dealings, for example local authorities. This is often cited as an objective for the design of HAs – for example the 2006 White Paper *Our Health, our care, our say* [10] (which lays down the rationale for the reorganisation of the English National Health Service in 2006) notes that providing “for PCT [i.e. HA] boundaries to be the same as those of local authorities with social services responsibilities, which would make it easier to achieve better integration of health and social care" is a desirable feature of a geography. (This cannot be the only objective as far as the White Paper is concerned since the post-2006 PCT structure does differ from the local authority structure in some respects: the White Paper is not very explicit about why this should be so.)

O3. HAs should be of the same size, for some suitable definition of "size" (e.g. population, value of spend). Reorganisations of HAs typically take place because HAs are deemed to be either "too big" or "too small". This seems to have been a major consideration in the 2006 reorganisation of the English NHS, which saw the number of HAs reduced from 303 to 152, and thus the size of the average PCT effectively doubled. Arguments for decreasing size often centre around, for example, the need of a PCT to be responsive to local needs, as large bodies are generally regarded as being excessively remote and bureaucratic; arguments for increasing size centre around the need for PCTs to reap economies of scale and pool financial risks – see Ref. [5] for an overview of competing arguments. Thus, planning tends to be predicated on the notion of an optimal HA size and deviations from that optimal size are likely to be frowned on (although identifying that optimal size is itself a multicriteria decision, based on weighing arguments for increasing against arguments for decreasing size). Moreover, different HA functions may have different optimal scales – while even a small HA may have a sufficiently large number of childbirths per year and thus be able to cost-effectively maintain the expertise to commission maternity services, the volume of demand may not justify maintaining specialist expertise on gender reassignment surgery. Thus decisions about what the statutory responsibilities of HAs are to be, and what responsibilities are to be reserved at the central level, are problematic if HAs differ markedly in size.

O4. HAs should be broadly comparable in terms of population characteristics which drive morbidity or cost of service for purposes of performance measurement and resource allocation. From an administrative point of view, a major difficulty in resource allocation to HAs is calculating the amount which an HA "needs" to provide services for a population with a given mix of population characteristics. A population which is disproportionately elderly presumably will consume more medical services than one which is disproportionately young, but how much? The statistical issues which arise in making this sort of quantitative needs assessment are extremely tricky, and results are notoriously sensitive to apparently innocuous technical assumptions [2]. Designing HAs so that they are as similar in population mix as possible would neatly sidestep this problem – at the extreme if all HAs have identical population mix and size, a fair allocation would be simply to give each PCT the same amount of money.

In the next section we will discuss how these ideas can be made operational in a way which helps in the design of a healthcare geography.

3. Framework for analysis

3.1. Formulation

Our approach to the issue under study will be a multi-objective programming one. A multi-objective program is a generalisation of a mathematical program or optimisation problem, which has several objective functions rather than one [11–13]. Because not all objectives can be simultaneously optimised, i.e., no single solution exists which is simultaneously at least as good or better on all objectives, the appropriate solution concept for a multi-objective program is the set of efficient solutions, i.e. those for which no solution exists which is simultaneously at least as good or better on all objectives.

First we describe the characteristics of the set of possible solutions. We define the elementary geographic units composing the territory of England as *wards*. We then represent the territory of interest as an undirected and planar graph called the contiguity graph $G(V, E)$, where $V = \{v_1, v_2, \ldots, v_n\}$ denotes the set of n nodes representing wards (elementary units) and $E = \{e_{ij}: i,j = 1,2,\ldots,n; i \neq j\}$ denotes the set of edges, where $e_{ij} = 1$ if the nodes v_i and v_j are connected and $e_{ij} = 0$ otherwise. Therefore, two wards are connected with an edge only if they are adjacent in the map.

Furthermore, we define the following sets of attributes as node weights:

- $P = \{p_1, p_2, \ldots, p_n\}$ denotes the set of populations of each ward.
- $A = \{a_1, a_2, \ldots, a_n\}$ denotes the set of areas of each ward in Km^2.
- $U = \{u_1, u_2, \ldots, u_n\}$ denotes set of numbers of unemployed (excluding economically inactive) of each ward.
- $O = \{o_1, o_2, \ldots, o_n\}$ denotes the set of numbers of people aged above 65 within each ward.

It is assumed that the desired number of HAs within the territory of interest has been determined and is denoted by N. Therefore, the task is to divide G by grouping its n nodes into N non-empty and disjoint HAs. A partition of G is denoted by $X = \{x_1, x_2, \ldots, x_N\}$, where x_i is a set of nodes corresponding to HA i. We then define the desired set of feasible partitions of G to satisfy the following constraints:

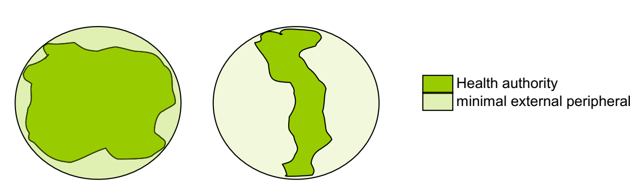
C1 Contiguity: the wards within each HA should be adjacent to each other i.e. within a HA there should not be a node disconnected from other nodes.

C2 Integrity: each ward should belong to one and only one HA. In other words, HAs should not overlap.

C3 Size of a HA: the number of wards within each HA should be within a defined range i.e. between some specified $L_{\min}$ and $L_{\max}$.

The objective functions which we will seek to optimise are as follows. Objective functions F1, F2, and F3 correspond to objectives O1, O2, and O3 outlined in the previous section. We operationalise O4 through two objective functions, F4 and F5, one of which is related to age and the other to economic activity.

F1 Compactness. Compactness reflects an aspiration that HAs should be closer to being ball-shared or circular. The idea is best understood graphically: see Fig. 2 for a representation of high and low compactness. Some sort of compactness concept features extensively in geographic partitioning and districting problems in social sciences and several methods for measuring compactness have been considered [14,15]; the measure we use is in line with common practice. Formally, supposing $AR(x_i)$ for $i = 1,2,\ldots,N$ denotes the total area of HA i (i.e. $\sum_{j \in x_i} a_j$) and $AR(\xi_i)$ for $i = 1,2,\ldots,N$ denotes the area of the wards whose centroids fall within the



a) High compactness *b) Low Compactness*

Fig. 2. Examples of high and low compactness.

smallest circle enclosing the centroids of all wards in HA i, we measure the compactness of HA i as:

$$C_i = \frac{AR(x_i)}{AR(\xi_i)} \tag{1}$$

Consequently, the measure of compactness for a HA takes a value in the interval (0,1). The objective function is therefore to minimise the average HA's deviation from the maximum compactness:

$$\text{Minimise :} \quad F1 = \frac{1}{N}\sum_{i=1}^{N}(1 - C_i) \tag{2}$$

F2 Co-extensiveness with Local Authority boundaries: It is desirable that HAs have common boundaries with a limited number of local authorities. This co-extensiveness presents advantages such as reducing administrative costs, creating shared knowledge and values, and more effective regional planning and decision making through the cooperation of local authorities with HAs [16]. We measure this co-extensiveness by measuring the degree to which a HA overlaps with different local authorities. To build the measure, rank the $m(i)$ local authorities in a HA i by the number of wards belonging to that local authority, in decreasing order. Call these local authorities $l_1 \ldots l_{m(i)}$. Define the number of wards in i belonging to local authority l_j as w_{ij}. Call the total number of wards in i, W_i. Define $M(x_i) = \sum_{j=1,\ldots,m(i)} j \times w_{ij}/W_i$ and construct the objective function as follows:

$$\text{Minimise :} \quad F2 = \frac{1}{N}\sum_{i=1}^{N}(M(x_i)) \tag{3}$$

F3 Size Homogeneity: The purpose of HAs is to make decisions on behalf of a local population, and as such, if they become too big, they become distant from that population. At the same time, because of fixed running costs, small HAs may not be financially viable. We incorporate the requirement that HAs should be neither too big or to small through this objective. We minimise the deviation of population size between HAs:

$$\text{Minimise :} \quad F3 = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{\overline{P} - P(x_i)}{\overline{P}}\right| \tag{4}$$

where the population of HA i is defined as $P(x_i) = \sum_{j \in x_i} p_j$ and the average population of HAs is defined as $\overline{P} = 1/N\sum_{i=1}^{N}(P(x_i))$.

F4 Population Age Homogeneity: We propose a measure for homogenising the population of people above 65 years old across HAs, by minimising the deviation of the ratio of Over 65's between HAs and formulate this objective as follows:

$$\text{Minimise :} \quad F4 = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{\overline{RO} - RO(x_i)}{\overline{RO}}\right| \tag{5}$$

where the number of over 65's in HA i is $O(x_i) = \sum_{j \in x_i} o_j$, the rate of over 65's in i is $RO(x_i) = O(x_i)/P(x_i)$ and the average rate of over 65's across HAs is $\overline{RO} = 1/N\sum_{i=1}^{N}(RO(x_i))$.

F5 Population Economic Homogeneity: In this study, we choose unemployment as the variable representing the deprivation level and aim to homogenise the HAs with respect to unemployment by minimising the deviation of their ratio of unemployed people (f_5) between HAs:

$$\text{Minimise :} \quad F5 = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{\overline{RU} - RU(x_i)}{\overline{RU}}\right| \tag{6}$$

where the number of unemployed people in HA i is $U(x_i) = \sum_{j \in x_i} u_j$, the unemployment rate in i is $RU(x_i) = U(x_i)/P(x_i)$, and the average unemployment rate across HAs is $\overline{RU} = 1/N\sum_{i=1}^{N}(RU(x_i))$.

The socio-economic data we use comes from the UK 2001 census. This data includes the Census Area Statistics (CAS) Ward level information on population, unemployment rates and number of people aged 65 and over for the electoral wards in the studied regions. This data has been obtained from NOMIS, a service of the Office of National Statistics (ONS).[1] Although somewhat out of date, the data is adequate for our current purposes which are to demonstrate the methodology and provide broad insight. The geographic spatial data on centroid coordinates, boundaries (used to compute the contiguity graph) and areas of studied electoral wards has been accessed through EDINA UKBORDERS and is available only for 2003 CAS wards. Due to the merger of a few wards in 2002, there are some differences between the datasets, and we have performed the reconciliation manually.

The objectives which we propose here are not uniquely compelling: for example, instead of (or as well as) co-extensiveness with local authorities, one could propose co-extensiveness with mental health trusts or ambulance trusts; instead of (or as well as) homogenising the elderly or unemployed population, one could homogenise population from a given socio-economic group, or the urban/rural mix. However, the objectives we have chosen are illustrative of the *type* of objectives which might be selected in a problem such as this. There is also a question of how aggregate or disaggregate objectives are: for example, economic and age population characteristics could be combined into a single index of need. However, our algorithm scales well with number of criteria and so we have taken advantage of this by decomposing the comparability objective O4 into two constituent objective functions.

3.2. Solution method

Solving multi-objective programs exactly is very challenging computationally and solving the problem outlined in the previous section exactly for a country the size of England is practically impossible with current optimisation technology. However, the computational difficulties associated with multi-objective programs have led to the development of various *meta-heuristic* methods with the aim of approximating and searching for efficient solutions which are both close to the true Pareto front and have a high diversity with respect to different objectives. These methods have a good track record of producing solutions, although the solutions do not come with optimality guarantees (as is the case with a classical optimisation method, such as the simplex method for linear programming).

The integer-coded multi-objective GA, proposed by Datta et al. [15] for partitioning a geographical territory into a given number of zones, is applied for solving the problem at our hand. It is a customised version of the multi-objective GA proposed in Ref. [15] as NSGA-II. NSGA-II is a general-purpose algorithm and it is customised here to work specifically for the partitioning problem only. The customisation is made by replacing the general-purpose operators of the original NSGA-II with special operators designed by incorporating the partitioning related information. To be specific, a solution of the GA is an array of n wards of the territory, where the value of an element of the array is the HA to which the representing

[1] http://www.nomisweb.co.uk/.

ward belongs. In order to speed up the search process by avoiding starting with an excessive number of infeasible (invalid) solutions, a greedy algorithm is applied here for initialising feasible (valid) or near feasible solutions by forcibly satisfying constraints as much possible. In this technique, a HA is first formed with a single random ward and then it is expanded to the neighbouring wards, which have not yet been included in any other HA. The expansion is continued until the permitted maximum size of the HA is obtained or all the wards are exhausted.

After the solutions have been initialised, the crowded tournament selection operator [17] is applied to the GA population of solutions. This operator selects two random solutions at a time, and a copy of the best one, based on the convergence and diversity of the solutions, is stored in a temporary population, known as a mating pool. A specially designed crossover operator for partitioning problem draws two random solutions from the mating pool and generates a new solution by inserting some random HAs from one parent solution into another. It also takes care of any overlapping, during this insertion, by relabelling the partially overlapped HAs as well as other HAs. Then, a partition-based mutation operator is applied to the solutions generated by the crossover operator. Since various objectives of our problem can be achieved only by balancing the sizes of the HAs, the mutation operator is engaged to alter the sizes of HAs by shifting a random outer ward of a HA to one of its boundary HAs.

We illustrate how the ideas of the crossover and mutation operators play out in the context of our problem with a small example (see Fig. 3.). In this example, we wish to partition a territory consisting of 21 wards into four HAs. To demonstrate the crossover operations, we have identified two solutions (chromosomes), which we call Parent 1 and Parent 2. We form a new solution by taking T3 from Parent 2 and inserting it into Parent 1, but this means that HAs P2–P4 are no longer viable (because T3 overlaps with all of them, violating the integrity constraint C2) and the solution has to be repaired, resulting in the creation of new HAs Q2 and Q3. In a similar way, another new solution can be generated from Parent 2 by inserting to it a HA from Parent 1.

As well as crossover, a mutation operator allows for a small chance of mutation. It searches through all small changes within chromosomes, and if a HA is selected for mutation, it will shift a boundary node of the HA i.e. increase the size of the HA by decreasing the size of its neighbour. This process will alter the size of the HAs to make solutions more diverse. Fig. 4 shows an illustration of mutation operation, where for the offspring 1 the boundary of P1 is shifted into Q3 and the newly created HAs are relabelled.

In order to make the GA converge faster, a problem-specific mechanism is applied for repairing an infeasible solution. Since an element of the considered solution array represents a ward of the territory and it is assigned only one value as the HA of the representing ward, the ward integrity constraint is automatically satisfied. If the HA contiguity constraint is violated under the proposed crossover operator, it is taken care of by a labelling mechanism by relabelling a disconnected portion of a HA as a new HA.

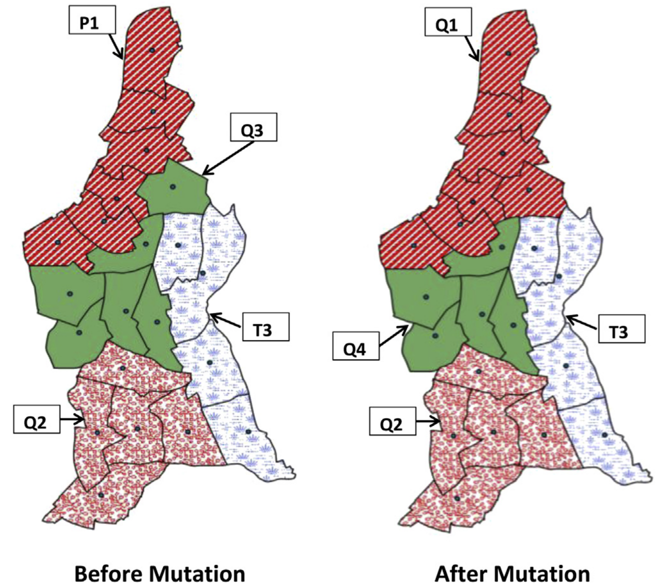


Fig. 4. An illustration of the mutation operation.

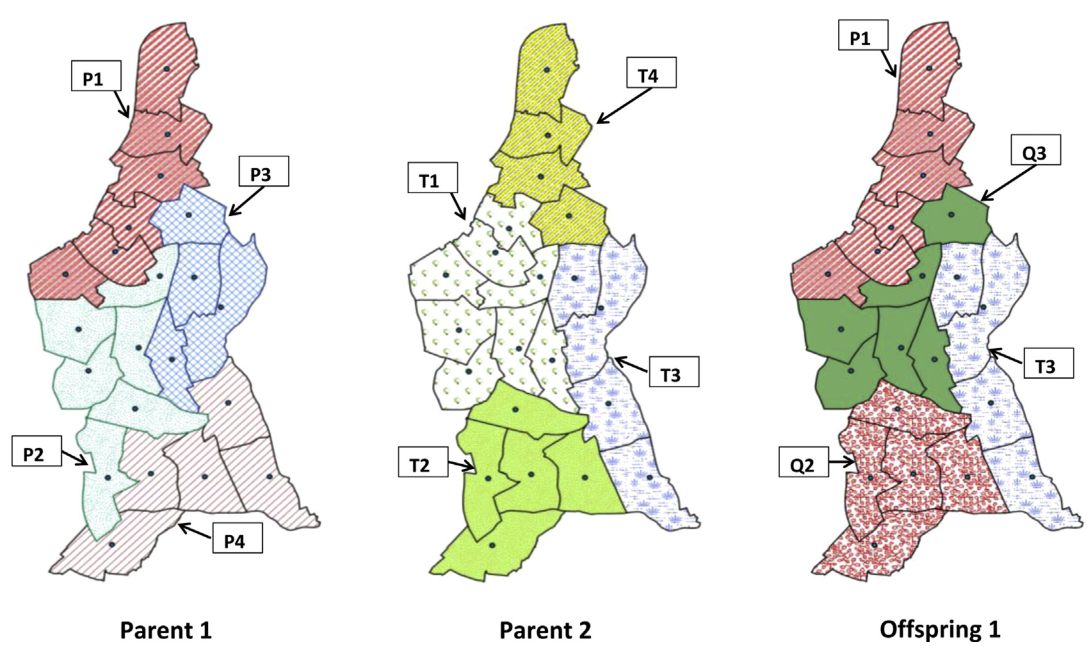


Fig. 3. An illustration of the crossover operation.

However, the HA size constraint may be violated at any stage of initialization/generation of a solution. Therefore, the repairing mechanism is applied here for steering an infeasible solution to the feasible region, as much as possible. Firstly it attempts to satisfy the minimum HA size, in which an undersized HA is eliminated by merging its wards in one or more adjacent HA(s). The process is continued until all such HAs are eliminated or the number of HAs reaches its lower limit. This process, however, involves the drawback of generating oversized HAs during the elimination of smaller HAs. Therefore, as the final step, the repairing mechanism attempts to reduce an oversized HA by merging some of its outer wards in adjacent HAs, if they have the capacity to take more wards.

If the above mechanism fails to repair an infeasible solution, the penalty-parameterless constraint handling approach [18], is applied to take care of an infeasible solution. It first makes an infeasible solution inferior to any feasible solution by assigning it a fitness value, then all the feasible and infeasible solutions are handled as feasible solutions only. Finally, the convergence and diversity based elite preserving mechanism [17] is applied in order to carry good solutions over generations. In this mechanism, both the parent and children populations of a generation are combined, and the combined solutions are sorted according to their quality measured in terms of their convergence and diversity. Then, the first 50% of the best solutions are extracted from the combined population for forming a new population for the next generation. This mechanism guarantees that, even if no good solution is generated at a generation, the GA never moves opposite to the optimum from the current position.

The proposed GA procedure is coded in C programming language. All computations are performed on a Dell Latitude 120L with an Intel® Celeron® chip (1.5 GHz) and 256 MB of memory, running on Linux Environment (Debian Release 5.0.4). The resulting maps are plotted by the ArcGIS 10 software.

4. Results and discussion

4.1. Characteristics of the East of England

Despite the effectiveness of the Genetic Algorithm in solving multi-objective problems, applying it to the health partitioning of the entire territory of England consisting of above 8000 wards would be computationally very demanding. For this reason, we focus on East England, which is one of the nine official government regions of England and also corresponds to East of England SHA. East England consists of:

- 6 Non-metropolitan counties: Norfolk, Cambridgeshire, Hertfordshire, Suffolk, Essex and Bedfordshire (which in turn contain 48 districts);
- 4 Unitary Authorities (UAs): Luton, Petersbrough, Thurrock and Southend-on-Sea.

While the counties cover large areas of mostly thinly populated countryside and have multiple districts, the UAs are smaller and contain highly dense cities and towns. At the lowest level of the subdivision hierarchy, there are 1118 electoral wards.

The territory of the East of England covers an area of 19,120 km^2 and has a population of about 5.6 million people. The density of population varies significantly across this region and is much higher in cities and towns than in rural areas and in the South than in the North. The density is highest in Luton Unitary Authority with 4400 people per km^2 and lowest in West Norfolk with about 100 people per km^2. This region has a higher proportion of people above 65 years of age (16.8%) than the England average (16.0%). The ratio of the population over 65 is much higher in north rural area and lower in the cities and southern region. Economically speaking, East England is one of the most prosperous regions in England with an overall high level of education and employment in comparison with the other regions. However this is not the case for all areas across the region and the unemployment rate (as a measure of deprivation) is generally higher in the North than South.

East England currently consists of 14 PCTs. Initial inspection of the PCT maps suggests that the current boundaries have been set with an emphasis on the co-extensiveness of PCTs to current local authority boundaries (although Great Yarmouth and Waveney crosses county boundaries and both South East and South West Sussex both contain a unitary authority and a slice of Sussex county), and to some degree on the compactness of HAs. However, at least at first glance, there is little evidence of taking size homogeneity and deprivation and health equality measures into account in setting the boundaries. Some extreme examples are Norfolk and Suffolk, where the high levels of compactness and coextensiveness with local authorities result in a high ratio of both unemployed and people over 65. Also, Peterborough and Luton have the same local authority and PCT areas, but this has been achieved at the expense of their low population sizes in comparison with the rest of PCTs.

4.2. Primary Care Trusts in the East of England

The first question which we examine is the efficiency of the existing partitioning of the territory of the East of England into 14 HAs ("Primary Care Trusts"). In answering this question, we took "local authority coextensiveness" to refer to co-extensiveness with the counties or unitary authorities of the East of England (there are 10 of these, and so they are on average roughly the same size as PCTs).

The final output of the algorithm is a set of solutions (partitions), which approximates the Pareto front, in which each solution is strictly better than others at least in one objective value. For conciseness, we will refer to such solutions as "efficient solutions". In general, it is hard to visualise these solutions, as to do so requires the visualisation of a points in 5-dimensional objective space. Hence, to illustrate some concepts, we present in Fig. 5, the results of a run for a two-dimensional subproblem using only objectives F1 and F3, compactness and size homogeneity. As can be seen from this figure, the set of efficient solutions forms an approximate Pareto front. This front may exhibit non-convexities in places, because even though a solution may be dominated by a convex combination of solutions, that convex combination may not be available because assignment of a ward to a Health Authority is a binary, zero-one choice. These sorts of non-convexities make this

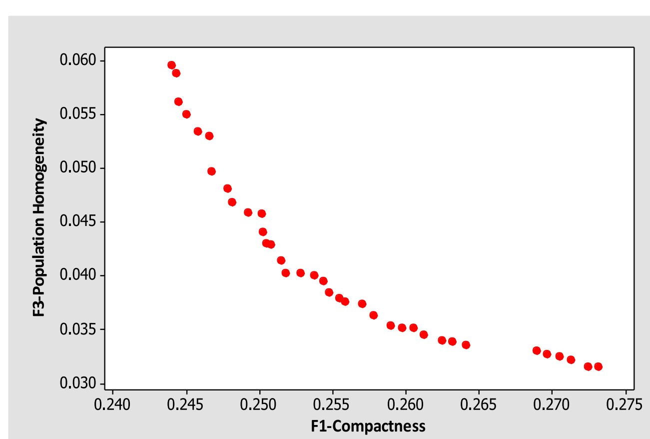


Fig. 5. Pareto front for a two-objective subproblem.

Table 1
Performance of modelled and actual solutions which minimise each objective function.

	Performance on objectives				
	F1	F2	F3	F4	F5
F1 Minimising solution	**0.193758**	1.249826	0.374486	0.134275	0.121036
F2 Minimising solution	0.339323	**1.057024**	0.434728	0.14088	0.141073
F3 Minimising solution	0.464833	1.252778	**0.011724**	0.118701	0.112224
F4 Minimising solution	0.540031	1.392056	0.389039	**0.046924**	0.155495
F5 Minimising solution	0.527363	1.255544	0.357812	0.135937	**0.039829**
Compromise	**0.419658**	**1.335425**	**0.271771**	**0.118175**	**0.117945**
Actual	**0.304887**	**1.082654**	**0.343147**	**0.13634**	**0.183611**

sort of problem computationally challenging and necessitate the methods which we use in this paper.

Turning now to our main run, we ran our algorithm for 10,000 generations of a population of 100 solutions, taking 7 h and 40 min. In Table 1 we display the performance of the efficient solutions which obtain the minimum values on each of the objectives (for the solution which minimises Fx, we have highlighted the score for that solution on Fx in bold), along with a compromise solution which secures good values on all objectives (but exceptional values on none). We also show the performance of the actual implemented partition of the East of England into Primary Care Trusts. It can be seen that the current implemented partition appears quite close to the solution which minimises F2, local authority coextensiveness. Indeed, the F2 minimising solution, although it is slightly better than the actual implemented solution on F2, only achieves this by worse performance on F3 (population homogeneity) and F4 (age homogeneity). So it appears, if we want to partition the East of England into 14 HAs, the current allocation seems to be a relatively efficient way of doing so, albeit one which places a relatively high weight on local authority coextensiveness.

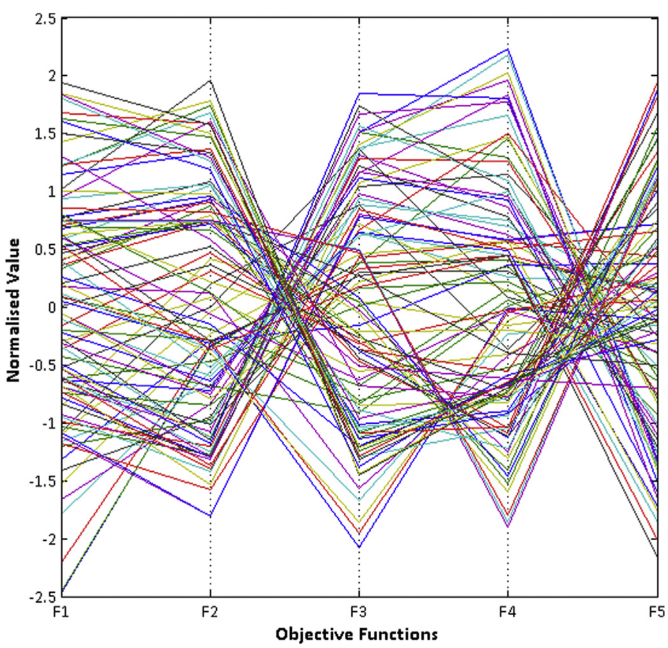


Fig. 6. Value path for efficient solutions.

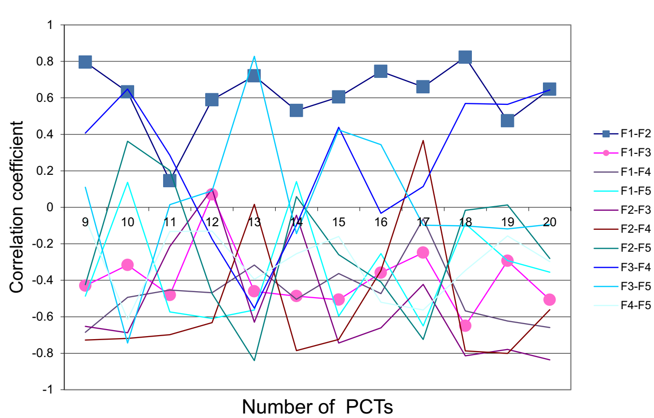


Fig. 7. Correlations between objective performance of efficient solutions for 10–20 PCTs in the East of England.

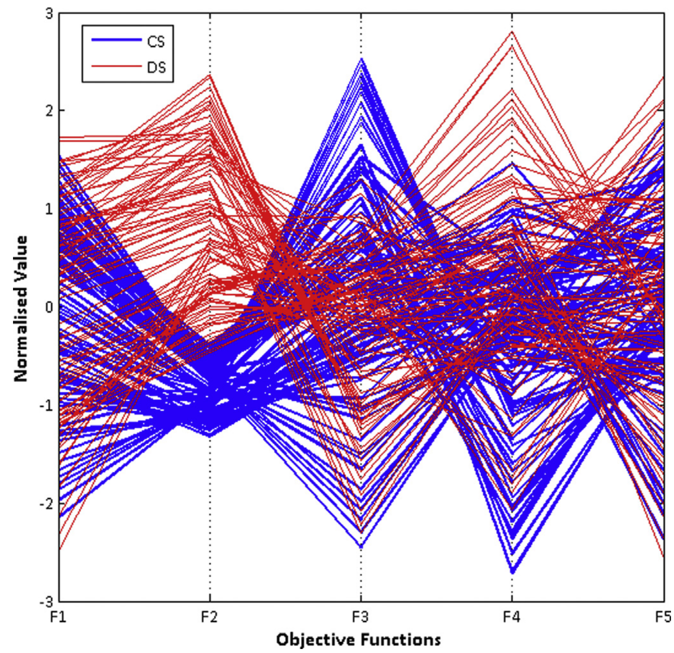


Fig. 8. Value path for 100 efficient solutions for CS and DS scenarios compare.

Fig. 6 shows the performance of each solution in the family of efficient solutions at termination as a so-called “value path” [11]. The value on the axes represents the spread of solutions for each objective and demonstrates the diversity of efficient solutions. The extent to which cross lines zigzag shows the sharpness[2] of the trade-off of objective values between different efficient solutions.

Recalling that the objective functions are to be minimised, it can be seen from Fig. 6 that efficient solutions which perform well on F1 (compactness) also perform well on F2 (local authority co-extensiveness). This makes intuitive sense: as local authorities are themselves compact, Primary Care Trusts which are coextensive with local authorities will also be compact. There appears to be a very sharp trade-off between co-extensiveness (F2) and size homogeneity (F3). Again this is not surprising as local authorities range in size by almost an order of magnitude: Essex has 1,310,836 people, while Thurrock has a population of 143,099. Similarly, good performance on F4 (age homogeneity) seems to be associated with

[2] What is meant by a “sharp” tradeoff here is a situation where good performance on one criterion can only be attained at the cost of poor performance on another criterion; a blunt tradeoff is one where relatively good performance on one criterion does not preclude relatively good performance on another criterion thus for a problem with two dimensions, {(1,0),(0,1)} is a solution set which exhibits a sharp tradeoff and {(1,0),(0.8,0.7),(0.7,0.8) (0,1)} exhibits a “blunt” tradeoff in this sense.

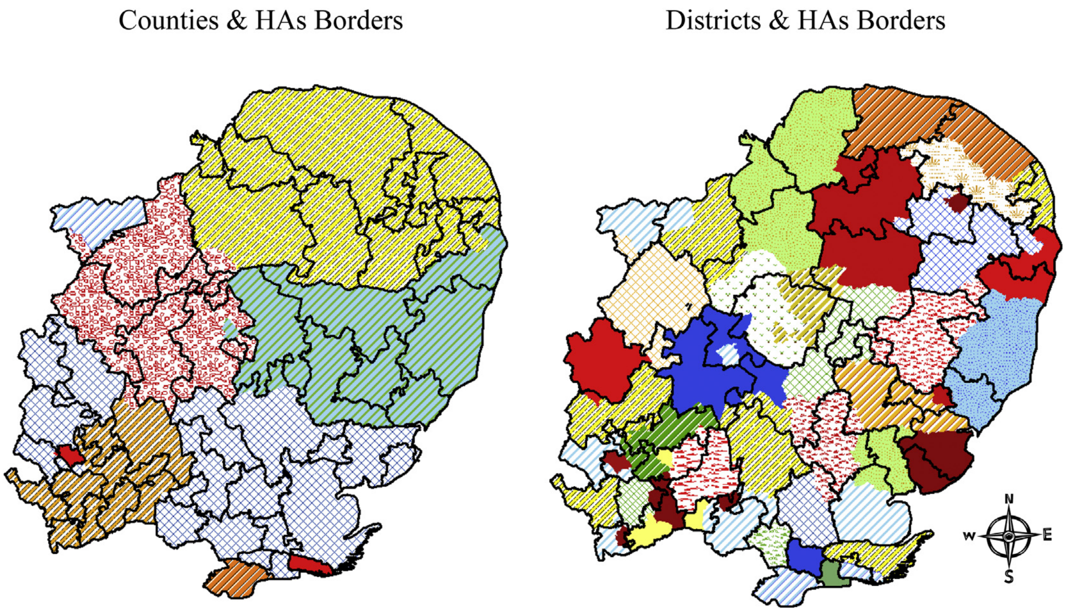


Fig. 9. Comparison between GP consortia and local authorities in East England.

good performance on F3 but to conflict with good performance on F5 (economic homogeneity).

A complementary perspective on the structural characteristics of the solutions is given by computing the correlation matrix for the solutions in the population at termination. We show in Fig. 7 the correlations between different objective values when we run our algorithm 12 times in order to partition East of England up into 9,10,11,… through to 20 PCTs. (For reasons of time constraints we run the algorithm for only 1000 generations and about 25 min each time, and a population of 50 solutions). It can be seen that the insights of the previous paragraph are borne out in this display: however many PCTs we partition the East of England into, there is always a strong positive correlation between performance on F1 *compactness* and F2 *local authority coextensiveness* (for convenience we have marked the data series for these correlations with squares); and there is almost always a negative correlation between performance on F1 and F3 (marked with circles), except when the number of PCTs is 12, when there is a slight positive correlation. Other patterns do exist in this dataset: for example it is striking that apart from the F1–F2 pair, there are no other pairs of criteria which consistently exhibit strong positive correlation. This suggests that there are indeed in general sharp tradeoffs between criteria, and validates our claim that a multi-objective approach is appropriate here.

4.3. GP consortia in the East of England

The second question we examine is how the East of England might be partitioned into smaller GP consortia. Although GP consortia are defined primarily by the participating GP practices, they nevertheless will have geographical responsibility, e.g. in order to ensure that someone has responsibility for the care of unregistered patients, and GP consortia have to identify the area for which they will be responsible in their constitution [19].[3] According to *GP Newspaper* [20] the size of the six pilot GP consortia in Cumbria is about 100,000. This is comparable with the size of the seven pathfinder GP consortia in the East of England, announced as this paper was in preparation, which have an average size of 146,000, according to the figures in Department of Health [21]. Working on this basis, there would be approximately 50 GP consortia in the East of England. We note that this is at the lower end of the scale suggested by the Department of Health [18],[4] but consortia are likely to be smaller in the East of England than the national average due to the predominantly rural nature of much of the region.

Co-extensiveness with local authorities remains important in this environment, and indeed has two possible interpretations: we could seek to make the GP consortia co-extensive either with the counties and unitary local authorities (the county scenario, CS), with the districts (the district scenario, DS). There are reasons why GP consortia might want to work collaboratively with either tier of local government: for example, counties deliver social care, which requires close coordination with healthcare delivery, but district councils have environmental and health and safety responsibilities.

We ran our algorithm for 10,000 generations of a population of 100 solutions. We compare the efficient solutions of CS and DS in order to observe the impact of local authority co-extensiveness level on the shape of our health authorities (Fig. 8). As would be expected from the construction of this index, the co-extensiveness with local authority objective function (F2) has a better value for county level than district level solutions (as there are fewer opportunities to violate co-extensiveness with counties than with districts). Solutions in the DS scenario which perform well on co-extensiveness with local authorities can be seen to perform badly on the age and unemployment homogeneity objectives (F4 and F5 respectively) relative to the CS scenario. At the same time many of the solutions under the DS scenario are amongst the best in terms of population homogeneity (F3) and seem also to perform quite well in terms of compactness (F1), so requiring co-extensiveness with districts has advantages and disadvantages in terms of the other criteria.

It is interesting to contrast the performance of the F2-extreme CS scenario (in which the objective F2 is co-extensiveness with counties) with the F2-extreme DS scenario (in which F2 is co-extensiveness with districts). These are shown in Fig. 9, with the

[3] Para 4.16 and 4.40.

[4] Para 4.21.

Table 2
Interobjective correlations for three different algorithm runs with different co-extensiveness definitions.

Objective pairs	DS	CS	HN
F1–F2	0.844	0.746	0.607
F1–F3	−0.585	−0.022	−0.493
F1–F4	−0.594	−0.364	−0.359
F1–F5	−0.536	−0.622	−0.633
F2–F3	−0.577	0.144	−0.607
F2–F4	−0.433	−0.402	−0.613
F2–F5	−0.764	−0.715	−0.836
F3–F4	0.284	−0.44	0.086
F3–F5	0.158	−0.303	0.39
F4–F5	−0.043	−0.054	0.328

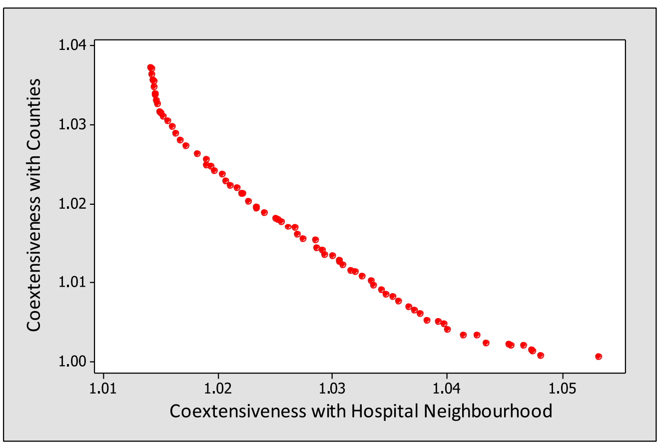


Fig. 10. Pareto front for bi-criteria run with two co-extensiveness criteria.

dark lines showing consortia boundaries and the coloured areas showing counties or local authorities respectively.

As can be seen, in the CS scenario our algorithm has performed well in securing a structure of consortia which are consistent with county boundaries, whereas in the DS scenario this has proved more difficult. Recalling that the algorithm run times are the same in both the CS and DS cases, this is reflective of the difficulty in securing a satisfactory allocation which is district-coextensive. A qualitative message to planners (and GPs) might be that if one constrains oneself with respect to the more granular district (as opposed to county) structure, the harder it will be to secure other objectives.

So far in our discussion of the co-extensiveness objective O2, we have supposed that the objective is to be co-extensive with local authority boundaries. However, decision makers may wish a partition into GP consortia to be coextensive with some other existing partition based, for example, on the location of acute trusts (so that GP consortia are able to build relationships knowledge of and relationships with a single, or small number of acute trusts). In order to demonstrate this, we computed, for each ward, the closest (minimum Euclidean distance) hospital. We call the resulting groupings of wards, “hospital neighbourhoods”, and consider them to be more appropriate than the traditional “hospital catchments” which are based on existing patient flows and so may reproduce current administrative arrangements rather than being reflective of natural patient flows. We re-ran our algorithm one further time in order to obtain a set of efficient solutions, using co-extensiveness with this hospital neighbourhood partition as our co-extensiveness objective (we call this the HN scenario). The inter-objective correlations of the efficient solutions for this run, as contrasted with the DS and CS runs, are shown in Table 2. It can be seen that the interobjective correlations for the HN run are quite similar to the interobjective correlations for the local authority runs (in no case is there an interobjective correlation in the HN run with opposite sign from the correlations in both the CS and DS runs), suggesting that the structure of the efficient set is quite similar.

Of course, rather than being motivated to trade a co-extensiveness objective off against other objectives, the decision maker may be interested in trading different sorts of co-extensiveness objectives off against each other. In order to explore this, we also performed an algorithm run (10,000 generations) which traded co-extensiveness with hospital neighbourhoods against co-extensiveness with county boundaries. Again we required 50 GP consortia, and allowed consortia to vary in size

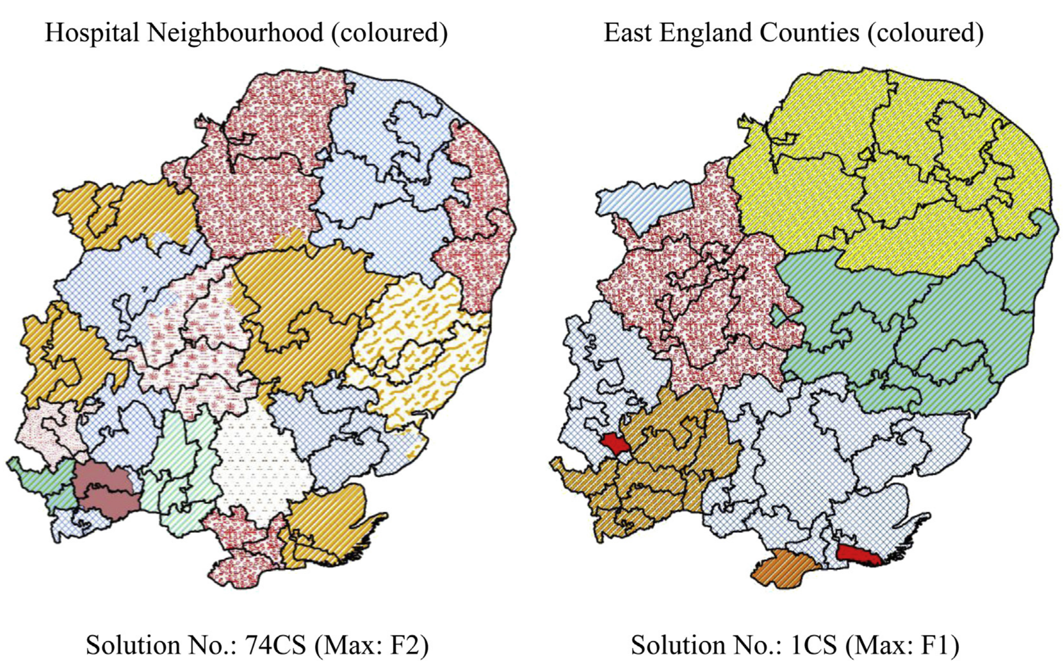


Fig. 11. Two extreme solutions for the bi-objective run.

between 3 and 200 wards in a consortium. The Pareto front for this bi-objective run is as shown in Fig. 10 and the extreme solutions which perform best on each of the objectives together with a good compromise solution, are shown in Fig. 11. The left-hand figure shows the extreme partition which performs best on co-extensiveness with counties, the righthand figure the extreme partition which performs best on co-extensiveness with hospital neighbourhoods. As in Fig. 9, the dark lines show consortia boundaries and the coloured areas show counties or hospital neighbourhoods respectively.

It can be seen from the Pareto front that although there are some candidate solutions which are highly coextensive with either the county boundaries, or the hospital neighbourhoods, there are no solutions which are coextensive with both. Indeed, the Pareto front appears to be fairly linear – there is no “elbow” which might naturally represent a good compromise solution between these two conflicting desiderata, underlining the need for the decision maker to think carefully about precise tradeoffs. The maps show where objectives might conflict – for example, a few wards in north-east Cambridgeshire are closer to a hospital in King’s Lynn in Norfolk than they are to any hospital in their own county.

5. Conclusion

In this paper, we have formulated the question of how to partition a territory into geographically defined commissioning units as a multi-objective program and solved this program using state-of-the-art optimisation techniques. The results are (in our view) insightful. One sample observation is that while tradeoffs between certain pairs of objectives (such as local authority co-extensiveness and compactness) may be rather blunt, tradeoffs between other pairs (such as local authority co-extensiveness and various sorts of homogeneity in size and population distribution) may be quite sharp. Another is that while in the past, Primary Care Trusts have been defined to be more or less contiguous with local authority boundaries, in the new liberated NHS, this constitutes less of a constraint on the formation of GP consortia (because of the smaller size of these entities), if local authority is interpreted to be county-level authority. At the same time, the statistical difficulties in performance measurement of and resource allocation to GP consortia, means that homogeneity of the consortia in terms of their morbidity characteristics, proxied in our model by their age and employment status, should loom larger. If one constrains GP consortia by requiring them to be coextensive with local authority districts, securing a partitioning which is attractive from that point of view will be that much harder.

The policy position is that GP consortia are to be formed “bottom up” from the grassroots. Nevertheless, there will inevitably by a role for the centre, whether the Department of Health or the new Commissioning Board, in ensuring that coverage is complete. It is very desirable that when the centre intervenes, it does so on a principled basis, otherwise decisions are will appear unfair, and may be open to challenge. Models like the one proposed in this paper can have a role in ensuring that decisions are made on a consistent basis across the whole of England.

Although our immediate motivation is policy situation in England, the sort of problem which we analyse is not specific to this country. Geographically defined entities play a prominent role in all Beveridge systems. The question of why these entities are defined as they are is rarely asked. When it is asked, it is typically answered in rather unsatisfactory ways, based on political whim or historical precedent. We show that there is a better way, and these questions can be brought within the orbit of science.

We also think there are still broader implications, in that this style of analysis (in particular the use of multi-objective techniques) seems to be underused in the health policy and healthcare management arenas. One of the features of healthcare is that system objectives are typically contested and particular actors may focus on one system objective to the exclusion of all others, or on one objective particularly intensely for a particular period of time. Forms of analysis which explicitly acknowledge multiple objectives seem to us to have great potential in sparking more consensual, reflective, and considered decision making.

Acknowledgements

This work uses data provided through EDINA UKBORDERS with the support of the ESRC and JISC. Boundary data used is from National Statistics/Ordnance Survey and is ©Crown Copyright 2006. Extracts are Crown Copyright and may only be reproduced by permission. Socio-economic and census data are made available by the Office for National Statistics through the NOMIS website. Maps were generated using the ARCGIS software. The work is carried out partially under DD’s post-doctoral research grant (SFRH/BPD/34482/2006) offered by FCT, the Portugese Science Foundation. AM and JF are grateful to the British Council and Council of Rectors of Portugese Universities for a travel grant, and AM would like to thank the Instituto Superior Técnico in Lisbon for their hospitality over several visits. JF and AM partially benefited from the LORIA grant “Multiple Criteria in ROAD”. The authors are grateful to Gwyn Bevan, Zack Cooper, Steve Gibbons and Jenifer Smith for helpful conversations in the development of this paper, and to the discussant (Rita Santos) and participants at the Winter 2011 HESG workshop in York for helpful and constructive comments on a preliminary version. The authors have no conflicts of interest.

References

[1] Edwards N. The triumph of hope over experience. London: NHS Confederation; 2010.
[2] Bevan G. The search for a proportionate care law by formula funding in the English NHS. Financial Accountability & Management 2009;25:391–410.
[3] Smith PC. Formula funding of public services: an economic analysis. Oxford Review of Economic Policy 2003;19:301.
[4] Bachmann MO, Bevan G. Determining the size of a total purchasing site to manage the financial risks of rare costly referrals: computer simulation model. British Medical Journal 1996;313:1054–7.
[5] Bojke C, Gravelle H, Wilkin D. Is bigger better for primary care groups and trusts? British Medical Journal 2001;322:599.
[6] Martin S, Rice N, Smith PC. Risk and the general practitioner budget holder. Social Science & Medicine 1998;47:1547–54.
[7] Department of Health. Welcome to the Department of Health. http://www.dh.gov.uk [accessed 08.09.09].
[8] Department of Health. Equity and excellence: liberating the NHS. London: Department of Health; 2010.
[9] Department of Health. Liberating the NHS: commissioning for patients – a consultation on proposals. London: Department of Health; 2010.
[10] Health DoOur health, our care, our say: a new direction for community services. London: The Stationery Office; 2006.
[11] Deb K. Multi-objective optimization using evolutionary algorithms. Chichester: Wiley; 2001.
[12] Ehrgott M. Multicriteria optimization. Berlin: Springer Verlag; 2005.
[13] Figueira J, Greco S, Ehrgott M. Multiple criteria decision analysis: state of the art surveys. New York: Springer Verlag; 2005.
[14] Vanegas P, Cattrysse D, Van Orshoven J. Compactness in spatial decision support: a literature review. Computational Science and Its Applications – ICCSA 2010;2010:414–29.
[15] Datta D, Malczewski J, Figueira JR. Spatial aggregation and compactness of census areas with a multiobjective genetic algorithm: a case study in Canada. Environment and Planning – Part B 2012;39:376–92.
[16] IDEA. Local authority and primary care trust partnerships. http://www.idea.gov.uk/idk/core/page.do?pageId=7314811 [accessed 26.10.10].
[17] Deb K, Pratap A, Agarwal S, Meyarivan T. A fast and elitist multiobjective genetic algorithm: NSGA-II. Evolutionary Computation, IEEE Transactions on 2002;6:182–97.
[18] Deb K. An efficient constraint handling method for genetic algorithms. Computer Methods in Applied Mechanics and Engineering 2000;186:311–38.
[19] Department of Health. Liberating the NHS: legislative framework and next steps. London: Department of Health; 2010.

[20] GP Newspaper. GP consortia – how white paper model can work 19 August 2010 [accessed 18.11.2010].
[21] Department of Health. http://www.dh.gov.uk/en/Aboutus/Features/DH_122384#_1 [accessed 07.01.11].

Dilip Datta is an Associate Professor in Mechanical Engineering at Tezpur University, India. Earlier he was a faculty member in Mechanical Engineering at National Institute of Technology Silchar, India, for sixteen years. He teaches mainly optimisation and other mechanical design based subjects. Datta did his graduation in Mechanical Engineering in 1992 from Gauhati University, post-graduation in Applied Mechanics in 1998 from IIT-Delhi and PhD in optimisation in 2007 from IIT-Kanpur. Datta obtained post-doctoral fellowship from FCT, Portugal, in 2008–2010. His research area is optimisation, especially population-based nontraditional techniques for multi-objective combinatorial optimisation problems.

José Rui Figueira is Associate Professor at CEG-IST, Instituto Superior Técnico, Technical University of Lisbon and Associate Researcher at LORIA Laboratory, Nancy, France. He also taught and did research at University of Évora, the University of Coimbra, and the Ecole des Mines de Nancy. He is a former member of LAMSADE, INESC-Coimbra and DIMACS research centres. He obtained his Ph.D. in 1996 from University of Paris-Dauphine and his Habilitation at the same University in 2009 in the field of Multiple Criteria in Operations Research and Decision Aiding. His current research interests are in decision analysis, integer programming, network flows and multiple criteria decision aiding, medical decision-making.

Arash Mostajeran Gourtani received the BSc degree in Statistics and Economics from University College London, UK, in 2009, the MSc degree in Operational Research from London School of Economics, UK, in 2010, and is currently pursuing his Ph.D. degree at University of Southampton. His research interests include multi-objective optimisation and stochastic programming with application to energy markets.

Dr Alec Morton is Senior Lecturer in Management Science within the Management Science Group at the London School of Economics. In 2008 he was seconded to the National Audit Office as Visiting Senior Fellow. He is interested in the application of systematic methods and modelling to help organisations deal with situations where values are contested and there are significant uncertainties. He completed an MSc in Operational Research and a PhD in Management Science at the University of Strathclyde. He completed his BSc in Mathematics and Philosophy at the University of Manchester.