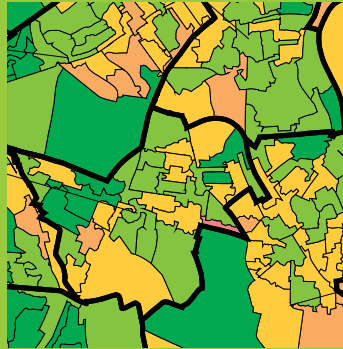


Predicting Housing Conditions



BRE

Every local authority housing officer needs to be able to target the worst areas. Robert Flynn explains how techniques developed by the Building Research Establishment can assist local authorities undertaking local house condition surveys and add value to existing data.

Part of the process of designing house condition survey contracts is explaining to councillors and senior officers how impossible it is to provide the ward data they covet with any accuracy unless huge and economically unjustifiable samples are selected. It is more typical for authorities to use smaller, less expensive samples and then divide the area into three or four groups of wards for reporting purposes. While this might be sounder statistically, it is, and remains, of little value for targeting purposes. Occasionally, areas where problems are suspected can be surveyed using larger samples but this is rarely a complete solution. Because they are so expensive, surveys cover several purposes and a sample, which is suitable for one purpose eg unfitness, may be unsuitable for another such as fuel poverty.

Recognising these problems, Chris Jarvis at the Greater London Authority, who advises London boroughs on local stock condition data gathering and Simon Nicol, director of the Building Research Establishment's Housing Centre had considered using English House Condition Survey (EHCS) data, which the BRE analyses for the Office of the Deputy Prime Minister, to produce a mathematical model to predict local housing conditions. In 2002, the BRE was successful in its bid to the Foundation for the Built Environment for a grant to fund the development of a local housing stock model.

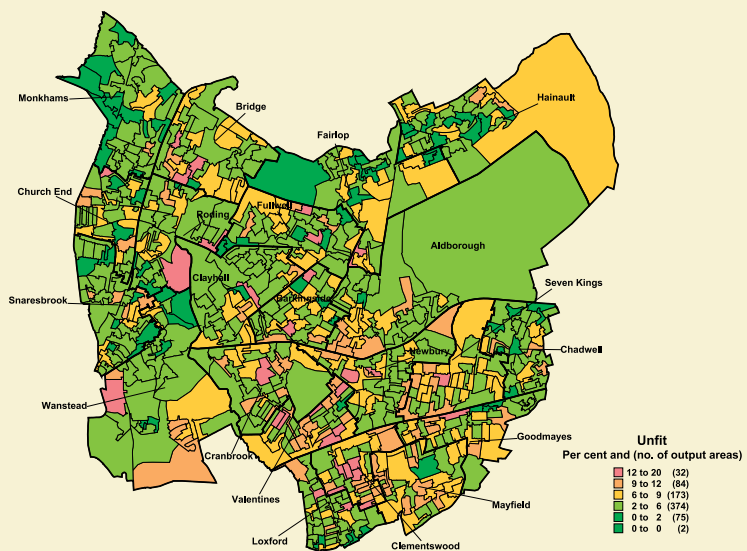
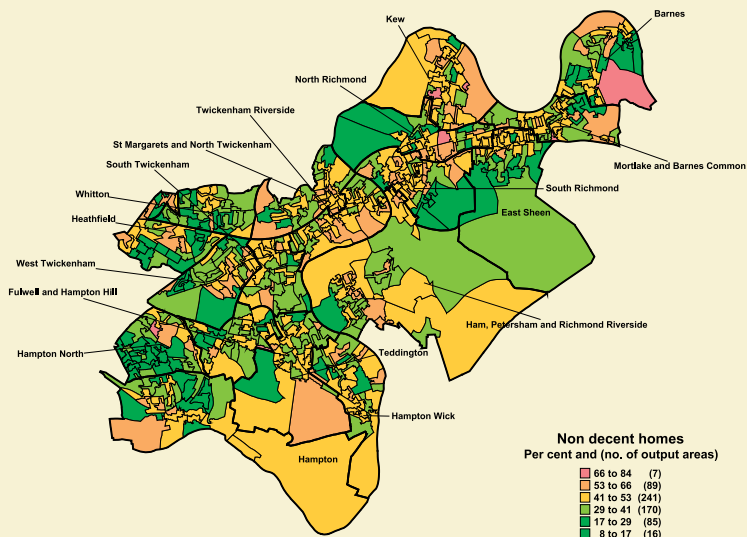
Eighteen months on, the BRE has designed a series of models which have taken EHCS 2001 data and combined it with data sources with national coverage (such as the census) to produce detailed maps for key indicators, notably non decent homes. BRE's initial development partner, the London Borough of Richmond provided invaluable feedback on ward data supplied by the preliminary models. This paved the way for the next stage in model development, which resulted in maps (see figures 1 and 2) that show typical outputs from the models for a local authority area (in this case non decent homes for the London Borough of Richmond and unfitness for the London Borough of Redbridge). Figures 1 and 2 show the percentage of non decent and unfit homes by census output area and statistical ward while box 1 (see over) provides a description of census output areas and statistical wards.

The full text of this article was originally published in the January edition of the Environmental Health Journal

So far, the BRE has produced models for:

- dwellings which fail the decent homes standard;
- dwellings which fail the decent homes standard due to unfit, inadequate thermal comfort, disrepair and non-modern facilities and services;
- non decent homes occupied by a vulnerable household;
- households in fuel poverty; and
- dwellings with a SAP rating less than 30.

Each model produces predictions of the percentages for each variable at the level of the local authority, the statistical ward and the census output area. The BRE was able to do this by using a method called CHAID, which is short for Chi-squared Automatic Interaction Detector.



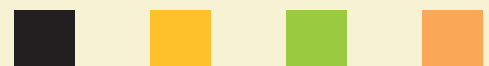
Census geography

Census output areas: these are now the smallest geographical unit the census reports on. They were intended to have similar population sizes and be as socially homogenous as possible (based on tenure of household and dwelling type), avoid urban/rural mixes and irregular shapes and tended to be constrained by obvious boundaries such as major roads. The specified minimum size was 40 households to ensure the confidentiality of data and the recommended size above 125 households.

Statistical wards: census output areas in England and Wales have been designed to fit into 2003 statistical wards, which reflect administrative boundaries promulgated (laid down in statute) by 31 December 2002. Most 2001 census outputs however, use Census Area Statistics (CAS) wards – these are a subset of 2003 statistical wards, with particularly small wards merged to protect data confidentiality.

The method as the BRE used it can be broken down into six stages:

- Select the dependent variable ie the EHCS statistic that is intended for modelling down to local level eg non decent homes.
- Prepare and establish links between national small area databases eg census and the EHCS data.
- Select the independent variables eg for non decent homes, any that may be related to the level of non decent homes such as dwelling type from the census, bearing in mind that one of the advantages of CHAID is it can accept large numbers of independent variables.
- Carry out the CHAID analysis. This is done using a computer programme called AnswerTree, which splits the database into groups that maximise the differences between them for each variable. This results in a diagram in the shape of an inverted tree with splits occurring at nodes (see figure 3). The independent variables that have greatest influence on the dependent variable are found at the top of the diagram. Once the differences become too small to be useful the programme stops splitting the database at what is known as a terminal node.
- The terminal nodes give predicted percentages for the dependent variable eg non decent homes, which can be assigned to every census output area in England.
- The percentages can then be summed to provide results to statistical ward and local authority level.



These six stages describe the process for the non decent homes model but each of the models was developed in a similar way.

While this may sound reasonably straightforward the process of preparing the data was a long and tortuous one of matching datasets that were difficult to reconcile. Until the 2001 census data and other related datasets arrived in August 2003, the BRE found it had been extraordinarily difficult to assemble the data in a usable format, which made it hard to establish whether the technique was yielding results of any real value. Now that the data has now been assembled and the models produce results, how reliable are they?

Before answering this question, it ought to be said that CHAID is considered to be a stage towards developing models rather than a tool, which produces a fully developed model. However, the results appear to be sufficiently encouraging to start making use of them.

The BRE has now supplied the models to two London authorities, Richmond and Redbridge, and their initial reactions have been very positive. Nevertheless, positive reactions are not proof of a model. The big problem has been and remains how can the models be proved when there is very little evidence to test against them? Furthermore, what should be regarded as proof?

To verify the models, the BRE has followed four approaches:

- Comparison with private sector house condition surveys. The results have been encouraging with survey data showing good agreement with the unfitness and disrepair models. The comparisons have, however, been limited to highly targeted surveys that yielded data for a few wards.

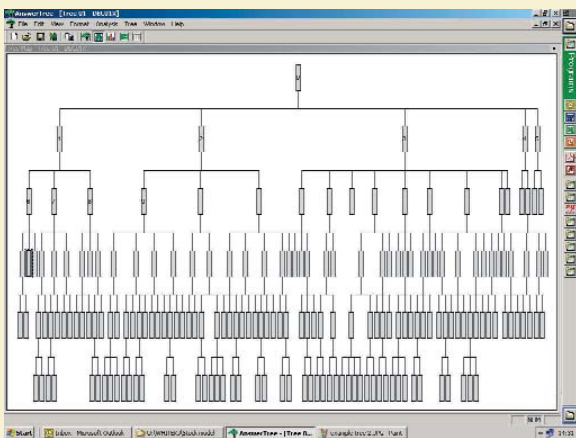


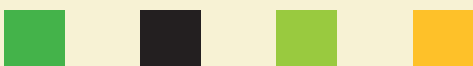
Figure 3: AnswerTree diagram illustrating how CHAID splits the data for non decent homes

Ward	Unfit (per cent)	
	Modelled 2001 data	1996 Survey
Clementswood	8	12
Valentines	8	12
Seven Kings	7	10
Mayfield	7	10
Newbury	6	10
Bridge	6	8
Cranbrook	6	8
Goodmayes	6	10
Aldborough	6	8
Fullwell	5	8
Clayhall	5	8
Barkingside	5	8
Church End	5	10
Wanstead	5	10
Roding	5	8
Loxford	5	12
Snaresbrook	5	8
Fairlop	5	8
Chadwell	4	8
Hainault	4	8
Monkhams	3	8
Borough	6	9

Table 1: Unfit dwellings: modelled and 1996 survey data

- Comparison with self-completion energy questionnaires. The problem of small samples is bound to dog any attempt at verification, so the large responses achieved by self completion energy questionnaires presented a means of verifying the SAP and fuel poverty models. However, in the one case where the BRE attempted this it found that problems of data quality and bias precluded their use at ward level or below.
- Comparison with data supplied by an energy company. The BRE used this information to develop a surrogate fuel poverty indicator, which had a correlation coefficient of 0.4 with the BRE's fuel poverty model. For the complex mix of social and physical data that make up the measure of fuel poverty the BRE felt that this was a good result at the level of the census output area. Certainly the fuel poverty model would have been of assistance in targeting the fuel poor in this local authority.
- Comparison with data supplied by specific local authority partners. The BRE was fortunate to be provided with good quality information by both of its local authority partners. As the data supplied by the London Borough of Redbridge provides the most recent comparison, it is important to consider this case in some detail.

The main statistic the London Borough of Redbridge provided was unfitness. Officers from the borough carried out high quality fieldwork for a 1996 survey, which was analysed by a team from the then London Research Centre. The wards were put into three groups for reporting purposes. The modelled data and the 1996 survey data are included in table 1.



There are four important points that should be made at this stage:

- there was a 3 per cent decline in unfitnes nationally between 1996 and 2001. If a similar decline took place in the London Borough of Redbridge then the modelled data is very close to the survey data;
- the modelled data tends to result in smoother data and therefore narrower ranges. The rank order is therefore of greater importance when making comparisons than the actual value predicted;
- the grouping imposed by the 1996 survey may mask real differences between wards within the groups; and
- the model includes all tenures whereas the 1996 survey excludes the social rented sector.

There are two very obvious areas of agreement between the model and the survey:

- two out of the three worst wards from the 1996 survey were predicted to be the worst and second worst wards by the model; and
- four out of the six wards in the second worst group of wards from the 1996 data were placed in the next six places.

This indicates a fairly good agreement between the model and the survey data at ward level although differences do exist.

The only other area comparison that could be made with the survey data was between substantial disrepair in the 1996 survey and the disrepair component of decent homes. While there are differences between these two measures they are sufficiently similar that a good agreement on rank order would be expected. In fact the disrepair model performed slightly better than the unfitnes model and there was close agreement with the survey data.

So far, users of the models have been less interested in the reports of the BRE's attempts at verification than in their own gut reactions to data provided for them. The maps in particular provide an easily assimilated impression of the predicted conditions, which result in a rapid and usually positive response.

As we near the end of the FBE-funded project, the BRE is considering how the knowledge gained might best be taken forward. The options include:

- seeking funding to develop the CHAID analysis into more formalised models using techniques such as logistic regression;
- seeking partners to provide good quality data to help verify the models;
- providing a service to authorities using the existing models; and
- developing models for other EHCS variables such as the housing health and safety rating.

The techniques described are primarily intended for use before a local survey has been carried out. This supports ODPM guidance [1], which promotes the ideal of using all available data sources to understand the local housing stock before commissioning a survey. Most of the BRE's efforts have therefore concentrated on developing a tool that will be of use before commissioning a survey. However, the question that is inevitably asked is, can this be used instead of a survey? This depends on what the purpose of the survey is. If all that is required is a strategic overview of conditions in an authority to inform a housing strategy, then this will probably suffice in the short term. The results could then be used to stratify future survey samples to target areas of interest in a much more focused way, possibly in the form of a rolling programme. Authority-wide results could still be achieved by including small samples in the areas of lesser interest.

This is not to say that the outputs of the models should be approached uncritically but neither should local data sources or the received wisdom on local conditions. The models merely add to a cocktail of information that needs to be considered in developing a strategy.

There is one other area that may prove to have as much potential as those already mentioned. A lot of good local surveys are undertaken that are only prevented on reporting small area statistics by their small sample sizes. Where the authority has confidence in its data, there is no reason why it cannot be used in place of the EHCS data to model down to census output area level. This is more time consuming than using the national model as the CHAID analysis has to be repeated, but if the local data quality is good then it should provide a more accurate picture of local conditions. This means that these techniques are of potential interest not only to authorities about to embark on surveys but also those looking to gain added value from their existing data.

Reference

- 1 Collecting, managing and using housing stock information: a good practice guide, ODPM 2000.

Further information

The work reported in this article was supported by the Foundation for the Built Environment. The author also gratefully acknowledges the part played by officers at the London Borough of Richmond and the London Borough of Redbridge in the development of the work. I would also like to thank Kevin White and Alan O'Dell at the BRE for their advice and assistance.

About the author

Robert Flynn is a principal consultant in the Housing Centre at BRE. He can be contacted at flynnr@bre.co.uk