

Estimating the number of homes in Scotland with internal lead piping

Gail Robertson, Amy Wilson, Yiannis Papastathopoulos, Margaret Graham, Lorna Eades, Efthalia Chatzisymeon

Project Summary

The aim of this project was to build a statistical model to provide an estimate of the number of houses in Scotland that are likely to contain lead piping or storage tanks and to identify postcodes which are likely to have increased numbers of houses with lead piping. This model would then be validated by carrying out two rounds of tap water sampling in postcodes throughout Scotland. To deliver this aim we first identified and collected appropriate datasets to be used in modelling, explored relationships between lead concentration and potential explanatory variables, made predictions of numbers of houses with lead piping within postcodes across Scotland using a best-fit model and validated the model using data collected in two rounds of tap water sampling.

We sourced datasets containing lead concentration values from tap water samples taken from houses across Scotland as well as variables expected to explain lead concentration in tap water, such as house type and house age from Scottish Water, the Scottish Government's Scottish House Condition Survey (SHCS) and other sources. The scope of the project was limited to domestic properties connected to public water supply systems across Scotland. The specific objectives of the project were to:

- Obtain data from Scottish Water and the Scottish Government containing relevant information on internal lead piping, water quality, communication pipe replacement and house characteristics.
- Develop a statistical model using available data sources to determine which characteristics of individual houses are associated with internal piping/storage and from this determine which postcodes are likely to have more houses with internal lead piping.
- Examine model goodness-of-fit and use model estimates to develop tap water lead concentration sampling protocol for houses within 600 postcodes.
- Use tap water survey results to test accuracy of model predictions.
- Provide an estimate of number of houses with lead piping/storage tanks in Scotland and identify street postcodes with greater numbers of houses with lead piping.

Results: Our model showed that house age and whether or not lead communication pipes within a postcode had been replaced were important variables for predicting whether houses were likely to have internal lead piping. Applying our model to house age and communication pipe data throughout Scotland, we identified postcodes which are likely to have high numbers of houses with lead pipes. To validate our model we received tap water samples from a stratified sample of 326 postcodes and found that our model was able to predict the percentage of houses with lead pipes accurately for postcodes predicted to have large number of lead pipes, but provided less accurate estimates of percentages of houses with lead pipes in postcodes predicted to have low number of houses with lead pipes and postcodes in the Glasgow area.

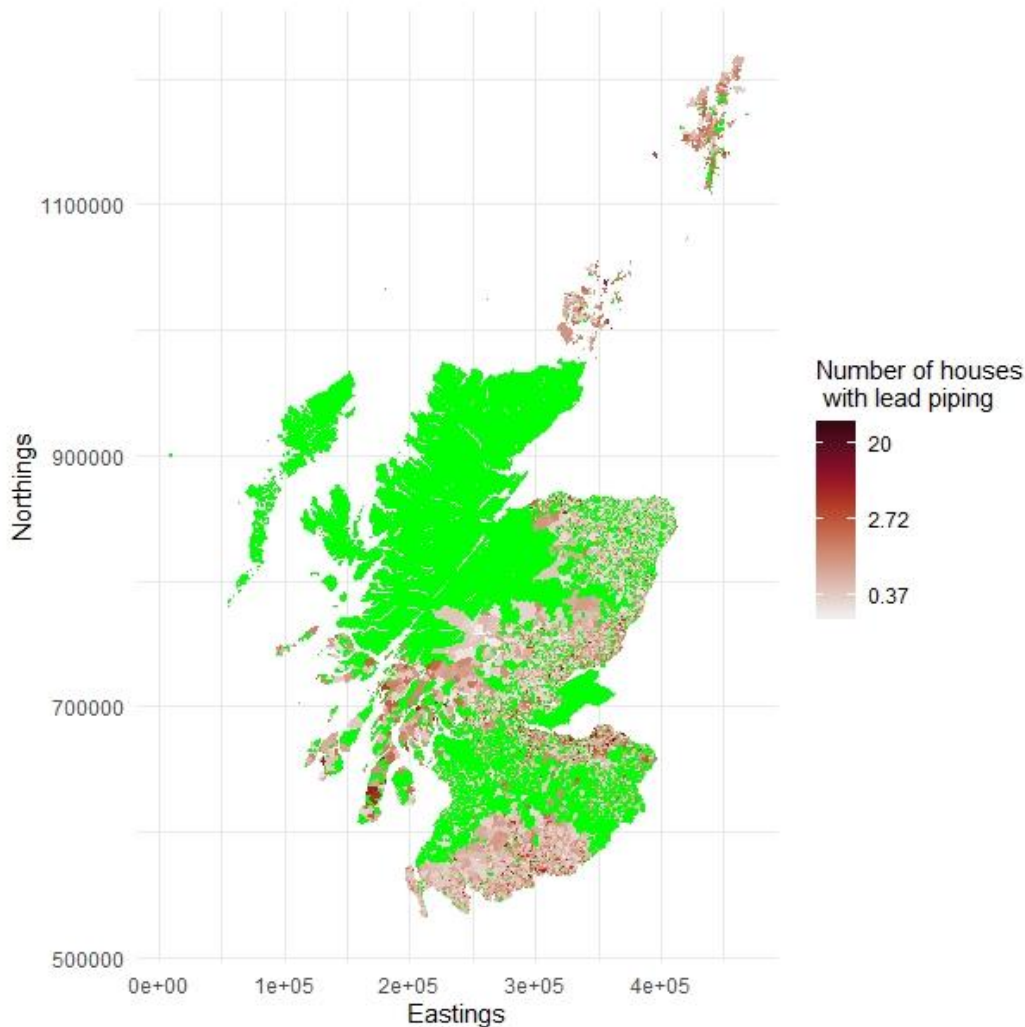


Figure 1. Map showing predicted numbers of houses with $\geq 1 \mu\text{g/l}$ for 70,180 street postcodes including house age data. Green represents areas with no explanatory variable data

Using our model to extrapolate the results across Scotland, our best estimate for the total number of houses in Scotland with lead piping was between 273,751 and 264,532 households (this varies depending on whether we include all postcodes with age data or only those in which $\geq 10\%$ houses have age data). We also estimate that the maximum number of households with lead piping (according to the upper level of a 95% prediction interval given by the model) lies between 1,397,912 and 1,232,554. Making predictions for all postcodes with explanatory variable data may lead to some inaccurate estimates if only a small proportion of houses in the postcode had age data. Only making predictions of number of houses with lead piping in postcodes in which more than 10% of houses had an age estimate leads to more accurate estimates of numbers of houses with lead pipes. Hence our estimate of 264,532 houses with lead pipes in Scotland is likely to be more accurate (with upper range of 1,232,554 according to 95% prediction intervals) (Figure 1).

Key recommendations:

- Data: The model developed in this study was based on data that did not cover all postcodes in Scotland. In some postcodes the data available was limited and potentially biased (e.g. by the method of collection). There are other possible variables that may impact the probability

that a household has lead pipes which could not be investigated due to lack of data. One key recommendation for improving the model and estimates described above is therefore to develop a comprehensive sampling scheme based on stratified sampling to ensure better coverage of samples across different postcodes and types of household.

- Statistical modelling: Improvements could be made to the model itself that may increase its accuracy. One such improvement is to account for spatial autocorrelation in the data. Another is to better capture uncertainty arising from the limited data.
- Validation: Model validation should be carried out using further tap water samples taken from a larger number of postcodes across Scotland, particularly those predicted to have lower percentage of houses with lead pipes. Currently model validation is based on tap water samples taken from 326 houses.

Project objectives (as specified in tender):

- i) Identify appropriate data sources and data sets to ensure a high degree of confidence across housing types and geographic location in the study
 - Obtained data from Scottish Water and Scottish Government surveys to explore variables affecting internal lead piping.
- ii) Utilise the appropriate datasets to develop a desk top model and establish the number of houses where lead may exist in the internal house plumbing system.
 - Used data to create model predicting areas containing more houses with lead piping
 - Made estimate of total houses with lead piping in Scotland
- iii) Verify the accuracy of the model through focused consumer surveys and water quality analysis and establish the level of confidence of the resulting data
 - Sent out sampling kits to further 600 customer addresses and received 326 responses
 - Model validation showed that model predictions were more accurate for postcodes estimated to have high numbers of houses with lead pipes, but underestimated numbers of houses with lead in areas with low numbers of houses with lead pipes and in the Glasgow area
- iv) The model should be able to identify the District Meter Areas (DMA) with greater prevalence of lead in houses. The use of unique property reference numbers will enable further granularity of data within the DMA.
 - DMA shapefile provided by Scottish Water showed that DMAs did not cover all of Scotland, so we made predictions at the street postcode level instead so that we could make predictions for as large an area as possible across Scotland.
- v) The model should be able to provide the basis of 'Domestic distribution systems risk assessment' required in the recast Drinking Water Directive.
 - We provide a map of postcodes across Scotland which highlight areas which are predicted to have greater numbers of houses with lead values $>1 \mu\text{g/l}$