Predicting population dental disease experience at a small area level using Census and health service data

Martin Tickle, Elizabeth Kay, Helen Worthington and Anthony Blinkhorn

Abstract

Background Information on the dental disease patterns of child populations is required at a small area level. At present, this can be provided only by expensive whole population surveys. The aim of this study was to evaluate the ability of Census data combined with health service information to provide estimates of population dental disease experience at the small area level.

Method Clinical dental data were collected from a large cross-sectional survey of 5-year-old children. A preliminary series of bivariate linear regression analyses were undertaken at ward level with the mean number of decayed, missing or filled teeth per child (dmft) as the dependent variable, and the Census and health service and lifestyle variables suspected of having a strong relationship with dmft as independent variables. This was followed by fitting a multiple linear regression model using a stepwise procedure to include independent variables that explain most of the variability in the dependent variable dmft.

Results All deprivation indicators derived from the Census showed a highly significant (p<0.001) bivariate linear relationship with ward dmft. The Jarman deprivation score gave the highest $R^2$ value (0.45), but the Townsend index ($R^2 = 0.43$) and the single Census variable ‘percentage of households with no car’ ($R^2 = 0.42$) gave very similar results. The health and lifestyle indicators also showed highly significant (p<0.001) linear relationships with dmft. The $R^2$ values were generally much lower than the deprivation-related Census variables, with the exception of the percentage of residents who smoked ($R^2 = 0.42$). None of the health or lifestyle variables was included in the final dental disadvantage model. This model explained 51 per cent of the variability of ward dmft.

Conclusions The results demonstrate the strong relationship between dental decay and deprivation, and all of the commonly used measures of deprivation exhibited a similar performance. For this population of young children health and health services shelf data did not improve on the ability of deprivation-related Census variables to predict population dental caries experience at a small area level.

Keywords: dental caries, prediction, deprivation

Introduction

It has been a long-standing wish of public health dentists to be able to predict groups of children at risk of developing dental caries. In the United Kingdom this desire is fuelled by a pressing need to target dental preventive programmes because inequalities in oral health are increasing and water fluoridation has still not been widely adopted. Local information about dental disease is also needed as an aid to service planning. At present, such information is provided by population surveys, which are expensive. An ability to predict population disease experience from information that is readily available would greatly reduce the costs of preventing disease, as the more disease-prone could be effectively and efficiently targeted.

It is well known that dental caries has a very strong association with socio-economic status. One possible means of predicting the dental disease experience in localities is by using small area indicators of socio-economic status. These indicators have been shown to have a strong association with dental caries at the small area level, but may in themselves be insufficient to capture aspects of health-related behaviour that are necessary to accurately predict small area disease experience. Seymour and Steele in a review of the literature relating to the relationship between dental disease and heart disease, hypothesized that dental disease as an outcome indicator could be a surrogate variable that summarizes a range of health and lifestyle variables that are associated with heart disease. Conversely, it is possible that general health outcome variables might act as summary statistics for lifestyle-related factors that influence dental health. Therefore these general health indicators may also provide opportunities for predicting dental disease patterns at a small area level.
As well as holding Census-derived variables, health authorities in the United Kingdom also store population mortality, lifestyle and health service related data at the small area level. The hypothesis underpinning this paper is that this health information together with area measures of socio-economic status may act as predictors of dental disease experience. This study, therefore, set out to evaluate the ability of Census data combined with health service data stored by health authorities to provide estimates of population dental disease experience at the small area level.

Method

This study used data collected in the 30 districts of the North West Region of England during the 1995–1996 NHS dental survey of 5-year-old children. The children were examined in school by trained and calibrated examiners according to standardized national guidelines. These survey data were aggregated to electoral ward level and a dmft score was produced for each ward. The dmft index is a measure of disease experience and measures the average number of teeth per child made up of decayed (d), missing because of extraction resulting from decay (m) and filled (f) teeth (t).

The Census variables held by the health authorities at ward level were available from two sources: the Small Area Statistics (SAS) dataset produced by the North West Regional Office of the NHS Executive, and the Office of National Statistics (ONS).

Ward-level data produced by NHS health services in the North West Region used in this study related to:

1. Hospitalization: ratios for discharges (including maternity but excluding well babies) for both elective in-patient or day cases and for emergencies, by selected diagnosis, and by selected procedure.
2. Mortality: standardized mortality ratios (SMRs) for selected diseases were also available for 1994–1996. The spreadsheets were compiled from ONS mortality records, which contain diagnoses for the original underlying cause of death.
3. Births: ward-based data relating to low birthweight, births to single mothers and total number of births were also available for the period 1992–1996. The source of these data was the birth records collected and stored by ONS.
4. Lifestyle: ward-level supplementary data on lifestyle from a survey commissioned by the North West Regional Office. These included unemployment rates, reported rates for smoking and people giving up smoking, and rates for residents who reported drinking semi-skimmed milk. The unemployment data related to 1996–1997 and were presented as a percentage of the adult population aged 16+. These datasets were converted into Microsoft Access tables and attached to the dental dataset using the electoral ward code as the linking variable. A preliminary series of bivariate linear regression analyses were undertaken at ward level with dmft as the dependent variable, and the Census and health service variables suspected of having a strong relationship with dmft as independent variables. Jarman and Townsend deprivation indices were included, as were the single Census variables: unemployment; percentage of households with no car; percentage of residents in Social Class 4 or 5; percentage of residents with higher education qualifications; percentage of children resident in households with no earners; percentage of households lacking amenities; a variable related to population mobility and the percentage of residents from an Asian ethnic background. The Limiting Long Term Illness variable from the Census was also included.

Ward-level health-related variables examined in these bivariate analyses included: percentage of births with a birthweight <2500 g; percentage of births alone; rate of hospital admissions for asthma; SMR all causes; SMR all cancers; rate of emergency admissions to hospital; percentage of smokers; percentage of residents drinking semi-skimmed milk; per cent unemployed as provided by a survey commissioned in 1996.

After the bivariate regression analyses a multiple linear regression model was fitted to ward dmft as the dependent variable to find the best ‘dental disadvantage’ model. In total, 72 independent variables were examined. A multiple linear regression model was fitted using a stepwise method with a probability for inclusion of 0.05 and exclusion of 0.1 to select the independent health and Census variables that explained most of the variability in dmft. All regression analyses were weighted according to the number of NHS dental survey subjects in each ward.

Results

Table 1 summarizes the total population of 5-year-olds and the sample size of the 1995–1996 NHS survey in each of the 30 districts. Seven districts undertook whole population surveys of all children present in school on the day of the survey.

Bivariate regression analyses

The bivariate relationships between dmft and the deprivation indices and single Census variables at ward level are displayed in Table 2. All indicators showed a highly significant \( p < 0.001 \) linear relationship with ward dmft. The Jarman score gave the highest \( R^2 \) value, explaining 45 per cent of the variability of ward dmft. The Townsend index and the single Census variable ‘percentage of households with no car’ gave very similar results, explaining 43 per cent and 42 per cent of dmft variability, respectively. All Census-derived variables had positive B coefficients except the percentage of residents with higher education qualifications \( (B = -0.08) \). This demonstrated an inverse relationship: dmft decreased as the percentage of residents with higher qualifications increased. The percentage of residents from Asian ethnic background in each ward explained only 16 per cent of the variability of ward dmft.
The bivariate relationship between ward dmft and ward-based health indicators is shown in Table 3. Again, all indicators showed a highly significant (p<0.001) linear relationship with dmft. The $R^2$ values were in general much lower than the deprivation-related variables in Table 2. The Census variable rate of residents with limiting long-term illness explained 27 per cent of the variability of ward dmft and the percentage of births in each ward with a birthweight below 2500 g explained 22 per cent. The rate of emergency admissions to hospital also provided a high $R^2$ value of 0.29. The percentage of residents who smoked could account for 42 per cent of the variability of ward dmft. The other lifestyle variable, percentage of residents drinking semi-skimmed milk, increased as dmft decreased. Interestingly, unemployment measured in the 1996 survey could explain only 19 per cent of variability in ward dmft compared with 30 per cent being explained by Census unemployment data dating back to the 1991 Census (Table 2).

Multiple stepwise regression at ward level

The model resulting from the stepwise multiple regression analysis is presented in Table 4. Seven variables made up the final ‘dental disadvantage model’: the percentage of households with no car; the percentage of workers in manufacturing; the percentage of households with more than four children; the percentage of households with dependants (child or sick pensioner); the percentage unemployed as a percentage of all economically active residents (from the Census); the percentage of children living in households with no earners; and the percentage of persons employed in primary production. None of the health or lifestyle variables was included in the final disadvantage model, which explained 51 per cent of the variability of ward dmft. Within the model the percentage of households with no car was the most powerful predictor, explaining 42 per cent of the variability, with ward dmft increasing by 0.04 of a tooth for every percentage point rise.

### Table 1 North West Region NHS survey of 5-year-old children 1995–1996*

<table>
<thead>
<tr>
<th>District</th>
<th>Sampling frame</th>
<th>Sample size</th>
<th>Records unable to be georeferenced</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blackburn</td>
<td>3931</td>
<td>3461 (88.04)</td>
<td>51 (1.47)</td>
</tr>
<tr>
<td>Blackpool</td>
<td>3693</td>
<td>647 (17.52)</td>
<td>4 (0.62)</td>
</tr>
<tr>
<td>Bolton</td>
<td>3755</td>
<td>317 (8.44)</td>
<td>3 (0.95)</td>
</tr>
<tr>
<td>Burnley</td>
<td>3449</td>
<td>3038 (88.08)</td>
<td>27 (0.89)</td>
</tr>
<tr>
<td>Bury</td>
<td>2370</td>
<td>2134 (90.04)</td>
<td>44 (2.06)</td>
</tr>
<tr>
<td>Central Manchester</td>
<td>1964</td>
<td>265 (13.49)</td>
<td>24 (9.06)</td>
</tr>
<tr>
<td>Chester</td>
<td>2096</td>
<td>382 (18.21)</td>
<td>334 (87.43)</td>
</tr>
<tr>
<td>Chorley and South Ribble</td>
<td>2740</td>
<td>2520 (91.97)</td>
<td>8 (0.32)</td>
</tr>
<tr>
<td>Crewe</td>
<td>3925</td>
<td>290 (7.39)</td>
<td>9 (3.10)</td>
</tr>
<tr>
<td>Halton</td>
<td>2013</td>
<td>1646 (81.77)</td>
<td>10 (0.61)</td>
</tr>
<tr>
<td>Lancaster</td>
<td>1538</td>
<td>1352 (87.91)</td>
<td>25 (1.85)</td>
</tr>
<tr>
<td>Liverpool</td>
<td>6148</td>
<td>308 (5.01)</td>
<td>107 (17.47)</td>
</tr>
<tr>
<td>Macclesfield</td>
<td>1975</td>
<td>383 (19.39)</td>
<td>18 (4.70)</td>
</tr>
<tr>
<td>North Manchester</td>
<td>1977</td>
<td>265 (13.40)</td>
<td>57 (21.51)</td>
</tr>
<tr>
<td>Oldham</td>
<td>3291</td>
<td>2396 (72.87)</td>
<td>1888 (88.73)</td>
</tr>
<tr>
<td>Preston</td>
<td>1708</td>
<td>1596 (83.44)</td>
<td>15 (9.94)</td>
</tr>
<tr>
<td>Rochdale</td>
<td>2968</td>
<td>267 (9.00)</td>
<td>21 (7.87)</td>
</tr>
<tr>
<td>Salford</td>
<td>3245</td>
<td>392 (12.08)</td>
<td>2 (0.61)</td>
</tr>
<tr>
<td>South Cumbria</td>
<td>2015</td>
<td>297 (14.74)</td>
<td>1 (0.34)</td>
</tr>
<tr>
<td>South Manchester</td>
<td>2083</td>
<td>296 (14.21)</td>
<td>68 (22.97)</td>
</tr>
<tr>
<td>South Sefton</td>
<td>2535</td>
<td>276 (10.89)</td>
<td>46 (16.67)</td>
</tr>
<tr>
<td>Southport and Formby</td>
<td>1270</td>
<td>390 (30.71)</td>
<td>13 (3.33)</td>
</tr>
<tr>
<td>St Helens and Knowsley</td>
<td>4878</td>
<td>612 (12.55)</td>
<td>8 (1.31)</td>
</tr>
<tr>
<td>Stockport</td>
<td>3559</td>
<td>370 (10.40)</td>
<td>10 (2.70)</td>
</tr>
<tr>
<td>Tameside</td>
<td>3564</td>
<td>2360 (66.22)</td>
<td>1940 (82.20)</td>
</tr>
<tr>
<td>Trafford</td>
<td>2696</td>
<td>426 (15.80)</td>
<td>33 (7.75)</td>
</tr>
<tr>
<td>Warrington</td>
<td>2472</td>
<td>286 (11.57)</td>
<td>3 (1.05)</td>
</tr>
<tr>
<td>West Lancashire</td>
<td>1478</td>
<td>715 (48.38)</td>
<td>19 (2.66)</td>
</tr>
<tr>
<td>Wigan</td>
<td>4040</td>
<td>372 (9.21)</td>
<td>4 (1.08)</td>
</tr>
<tr>
<td>Wirral</td>
<td>4579</td>
<td>744 (16.25)</td>
<td>14 (1.88)</td>
</tr>
<tr>
<td>Total</td>
<td>87957</td>
<td>28805 (32.75)</td>
<td>4806 (16.68)</td>
</tr>
</tbody>
</table>

*Total population used as the sampling frame, sample sizes (number, with percentage of sampling frame given in parentheses) and records unable to be georeferenced by district (number, with percentage of sample size given in parentheses).

†Whole population surveys.
The Census unemployment variable from the SAS dataset was retained in the model, whereas the 1996 ward unemployment rate provided from survey-derived data was dropped.

**Discussion**

The NHS child dental health surveys are nationally co-ordinated by the British Association for the Study of Community Dentistry (BASCD). Since 1984 each health authority/board in the United Kingdom has commissioned a rolling programme of data collection, within which 5-year-old children are examined every 2 years. In 1992 BASCD established a national core protocol and introduced a training pack for examiners. In 1994 national sampling guidelines were introduced to improve representativity. To ensure continuity and the ability to permit comparison over time,
data are still collected according to the health authority boundaries of 1984. To satisfy the need for information at a small area level there has been a tendency to collect whole population data. In the study reported here, whole population surveys were performed in seven districts. Aggregation of these large datasets provides an opportunity to examine the relationship between dental disease and various health and socio-economic indicators at small area level.

All of the deprivation indices and single Census variables commonly used to measure disadvantage showed a significant linear relationship with dmft. The Jarman index provided the highest $R^2$ score (0.45), slightly higher than the Townsend index ($R^2 = 0.43$). The fact that the performance of these two indicators was so similar substantiates the view that these deprivation indicators have a strong level of agreement.14,15 The Jarman score contains a demographic variable based on the proportion of young children resident, which probably accounts for its marginally better performance than the other indicators for this population of 5-year-olds. It also adds credence to the notion that the inclusion of demographic variables makes deprivation indicators more robust.16

Interestingly, households with no car is one of the four variables that make up the Townsend index, and of the single Census variables it emerged as the best performer ($R^2 = 0.42$) in terms of explaining variability of dmft at ward level. Car ownership is seen as being a good proxy for income17 but it also has specific, associated health-damaging factors.18 For example, it could be hypothesized that lack of a car can mean that these households have poor access to shops selling healthy foods and therefore a detrimental effect on dietary patterns. When aggregated to small area level the percentage of households in Social Class 4 and 5 under-performed compared to the other indicators of deprivation. This could be because it is not designed to be an area measure of deprivation; more probably, the under-performance is due to the well-documented failings of occupational social class as an indicator of deprivation,19 resulting in its limited ability to accurately reflect inequities in health.20

Unemployment as recorded in the Census did not perform particularly well, explaining only 30 per cent of dmft. This was unexpected, as unemployment as a single variable has been advocated as a strong indicator of deprivation and is closely related to health status.14,17 The log transformation of unemployment is included in the Townsend index to reduce the skewness of the variable. Because of this and the poorer than expected performance of unemployment as a predictor, a log transformation of Census unemployment was performed and a bivariate regression analysis demonstrated a highly significant linear relationship with dmft ($p < 0.001$) but increased the $R^2$ statistic only marginally from 0.30 to 0.36. This low coefficient score may reflect the transient nature of unemployment,16 and with increasing employer demands for a more flexible workforce and the large inequalities in the earnings of the employed, it may be a less meaningful indicator of deprivation in the 1990s. Interestingly, the 1996 survey-based measure of unemployment explained only 19 per cent of the variability of dmft. This perhaps represents the fact that the Census measure was a percentage of all the economically active population, whereas the survey data measured the unemployed as a percentage of all adults aged 16 plus. It probably also reflects the greater spatial integrity of Census data. Children from an Asian ethnic background experience higher levels of tooth decay in the deciduous dentition than children from other ethnic backgrounds.21 However, the Census variable ‘percentage of residents of Asian ethnic background’ explained 16 per cent of dmft variability. This is probably attributable to the distribution of this variable, which will be highly skewed because the Asian population is relatively small and geographically clustered in specific areas of the Region.

The bivariate relationships with the various health variables showed that the percentage of residents who reported being smokers explained 42 per cent of ward dmft, as much as the percentage of households with no car. Although all of the other health-related variables demonstrated a significant linear relationship with dmft, the $R^2$ statistic scores were generally much lower than for the traditional Census measures of socio-economic status.

**Multivariate stepwise analyses using Census and health variables**

To try and improve on the performance of the deprivation indices and single Census variables a multivariate model was constructed. The variables in the final ‘dental disadvantage’ model all came from the Census and gave a marginally higher $R^2$ score (0.51) than the Jarman indicator (0.45) alone. The $R^2$ reported here were lower than those reported in Salford by Jones et al.5 using the Jarman index ($R^2 = 0.88$). This poorer performance probably results from the much broader geographical area covered in this study and the greater number of wards involved. Locality-specific effects across the Region will also overlie the relationship between deprivation and dental disease at ward level and therefore tend to reduce the $R^2$ statistic. The resultant dental disadvantage model is also likely to be data specific and these analyses would need to be repeated on different populations to determine if the same independent variables consistently appear in the disadvantage model.

None of the health-related variables were included in the deprivation model. This suggests that the area measures of general health outcomes did not provide an additional health-related behavioural dimension to improve the prediction of dental caries experience at ward level. This was disappointing, as the literature suggested a link would be found between dental disease and general ill-health, as a result of common health risk behaviours. Different reasons may exist for the lack of an independent effect of these health variables. Payne and Locker22 found significant but weak correlations between
general health and oral health risk behaviours at the individual level. It may be the case that these associations can be found at the individual level but the association may not be strong enough to show a significant relationship at the area level.

Also, many of the diseases represented in the NHS datasets are diseases of old age and possibly less appropriate to predicting disease experience in young children. However, they should offer a glimpse of the health profile of small areas, which one would have thought would be reflected in the dmft of the resident child population. Migration and time from risk exposure to disease outcome (death) may be responsible for the non-significant findings. Many of the diseases under consideration take many years to develop. Also the social and physical environment of areas can change and therefore the same risk factors for young children may no longer be present. As a result of these age and cohort effects these measures may be more useful in identifying areas with high levels of adult oral disease and its consequences. The Limiting Long Term Illness variable of the Census may be especially useful in this respect. If this were to be the case, it would be an extremely important finding, as local information on oral disease and its consequences in the adult population does not exist in the United Kingdom. Further research is needed into this question.

Not all of the health variables examined related to older adults. Lone birth rates were also positively associated with ward dmft, but lone births are also related to socio-economic status. Low-birthweight rates (<2500 g), a widely used indicator of the health of young children, was not a strong enough predictor of dmft to be included in the final deprivation model, as was the rate of lone births. This indicator was expected to offer more promise of helping to explain dmft at a small area level, but lone-birthweight rates are also related to deprivation. It would therefore seem that the strong relationships that general health variables have with deprivation ensured that significant, bivariate linear relationships with ward dmft were found, but once deprivation was controlled for in the multiple regression model, its confounding influence was exposed.

It may be more useful to look at the relationship between dental disease and general health from a different perspective. The NHS dental datasets can provide high-quality morbidity data at small area level and these data may be useful to public health medicine colleagues as indicators of the population health and risk-behaviour of children of school age; for example, for child nutrition and parental health-related behaviour. Further collaborative studies are needed to examine the relationship at a small area level between other indicators of the general health of school children and the regularly collected dental disease data.

In conclusion, the results demonstrate the strong relationship between dental decay and deprivation at a small area level and that all of the commonly used measures of deprivation show the same association. For this population of young children health and health services shelf data did not improve on the ability of deprivation-related Census variables to predict population dental caries experience at a small area level. However, there is still a need to be able to provide information on dental disease experience at a small area level cost effectively. The health indicators may be more effective in predicting population oral disease and its consequences in older populations. Further work is needed in this area.

Acknowledgements
The authors would like to express their thanks to the Community Dental Services of the districts involved, and the North West Dental Public Health Resource Centre.

References
16 Doll H, Mertens B, Kleinschmidt I, et al. A standardisation approach to


Accepted on 7 February 2000