Extending COPD Prevalence Modelling to Very Small Areas

Meic Goodyear

Public Health Intelligence Specialist

NHS South East London

Public Health Lewisham

3rd Floor, Laurence House,

1 Catford Road, SE6 4RU.

Telephone:0207 206 3273

email: meicgoodyear@nhs.net

Abstract

Background Numbers of cases of COPD known to general practitioners are substantially lower than the best modelled estimates of prevalence. More complete diagnosis in primary care should lead to better care in this long term condition.

Methods Age, gender and postcode of residence of all registered patients resident in a London borough were combined with smoking prevalence estimates and from the Integrated Household Survey, deprivation data from the IMD 2007 and ethnicity data from the 2001 census were used to populate a nationally-recognised model to derive prevalence odds ratio density at very small area levels.

Results The modelled prevalence of COPD varies considerably over small areas, some postcodes have more than 20 times the probability of a case than others within the same borough.

Conclusions This approach could form a basis for formulating effective case-finding strategies at GP practice level.

Introduction

Background

Nacul et al[1] have developed a model for expected prevalence of Chronic Obstructive Pulmonary Disease (COPD), which has been accepted and recommended by the Association of Public Health Observatories (APHO). NHS Lewisham used this model when comparing recorded and expected prevalence of COPD at General Practice level in its Annual Public Health Report for 2009[2].The ratio of recorded to expected prevalence in Lewisham was slightly above the average for London practices, but substantially below that for England. Nacul et al[3] have compared recorded and expected prevalences at Local Authority level across England. They conclude that there is a North-South gradient (higher recorded prevalence in the North), that ratios were lower in urban areas than in rural, and that there was a particular problem of under-diagnosis in London though highly mobile populations may contribute to this effect.

Their model uses multinomial logistic regression to arrive at odds ratios based on age, sex, ethnicity, rurality, smoking status, and deprivation scores, from which estimated numbers of people age >16 were calculated. This study examines the possibility of using the model, in conjunction with other available data, to map the expected risk at very small area level (to individual postcode) in order to support the targeted process of case-finding.

Method

Details of age, gender, and postcode of residence were obtained for all patients with a Lewisham GP and resident in the Borough. Patients were age-banded according to the grouping used in the model (<35, 10-year bands to 74, 75+).

Smoking status was not available below borough level, so will not vary across the borough in these calculations. The developers of the model obtained their estimates of smoking prevalence from the Health Survey for England (HSE) 2001[4]. There has been a considerable shift in attitudes to tobacco smoking in the intervening ten years, with sustained targeted stop smoking campaigns, supported by changes in taxation and legislative restriction of smoking in public places. For this exercise it was considered desirable to use more up-to-date estimates if possible. The *smoking prevalence estimates among adults (18+)* from the Integrated Household Survey 2009-10[5] were used. These have been preferred to the earlier modelled estimates as they are based on measurements

taken after the implementation in 2007 of the ban on smoking in enclosed public places. They are lower (for Lewisham) than the 2001 HSE estimates, resulting in lower expected prevalence numbers. Published data is not available below local authority level, but the IHS database includes some data items at lower geographical levels, and ONS made available data at electoral ward level. This data is classified as "experimental" statistics rather than official statistics, and as such may be subject to modification. To allow comparison with other sources, our results were then scaled up proportionally to agree with the overall total estimated numbers from the model. Smoking status is classified as Never smoked/Former smoker/Current smoker.

Output Area (OA) of residence and Lower Super Output Area (LSOA) of residence were derived from the NHS Postcodes Directory[6] for each patient's postcode of residence.

Ethnicity was estimated using 2001 Census data at OA level. The categories used in the model are White (including all mixed ethnicities and other groups, as defined in model documentation), Black, and Asian. While there are known to have been substantial demographic changes in the interim period, no more recent population estimates at OA level are believed to exist.

Deprivation was derived from the LSOA of residence, and quintile within the national distribution was calculated, using the Index of Multiple Deprivation 2007[7].

Urbanity was not included, as the whole of Lewisham is urban, and the odds ratio for urban area is 1.

From these categories, the odds ratio for each person for each of the factors was calculated. As odds ratios are multiplicative, an overall odds ratio for each person could be derived as the product of the individual odds ratios. In general, if a patient has a set of n odds ratios for a particular condition, the patient's overall odds ratio is

$$O_p = \prod_{i=1}^{n} O_i$$

The cumulative risk, defined as the total odds ratio associated with the area, was then derived as the sum of the odds ratios of persons resident within an area. If there are m persons in an area, the formula is

$$O_a = \sum_{j=1}^{m} O_{p,j}$$

The overall prevalence rate P was calculated as the expected total number of COPD patients (from the model) divided by the total number of resident persons in the age range.

The calculated risk population for an area C_a, was defined as

$$C_a = P * O_a$$

The expected number of COPD patients in an area, C_a, was calculated as

$$E_a = C_a * P$$

The risk per registered person in an area was calculated as

$$R_a = \frac{C_a}{m}$$

The results were mapped at postcode level in Manifold GIS[8], using boundary files from Codepoint with Boundaries[9] to plot COPD risk density across the borough.

Discussion

Main findings of this study

Figure 1 shows the risk density at individual postcode level (the sum of the odds ratios for all registered patients living in the postcode boundary). The shading can be interpreted that the chance of finding a COPD case in a postcode area with the darkest colour is more than twenty times that of finding a COPD patient in a postcode with the lightest shading, which suggests an efficient strategy for case-finding.

While some quite wide areas have low risk density, and others are overall high, in many areas postcodes having a high risk of COPD are very close to postcodes of low risk. Also, the pattern of risk does not directly reflect the overall deprivation in the borough (as measured by the IMD). Patients resident in parts of the extreme north and

extreme south of the borough, which have the highest deprivation, have relatively low risk of COPD, while those in the least deprived area are at high risk of COPD.

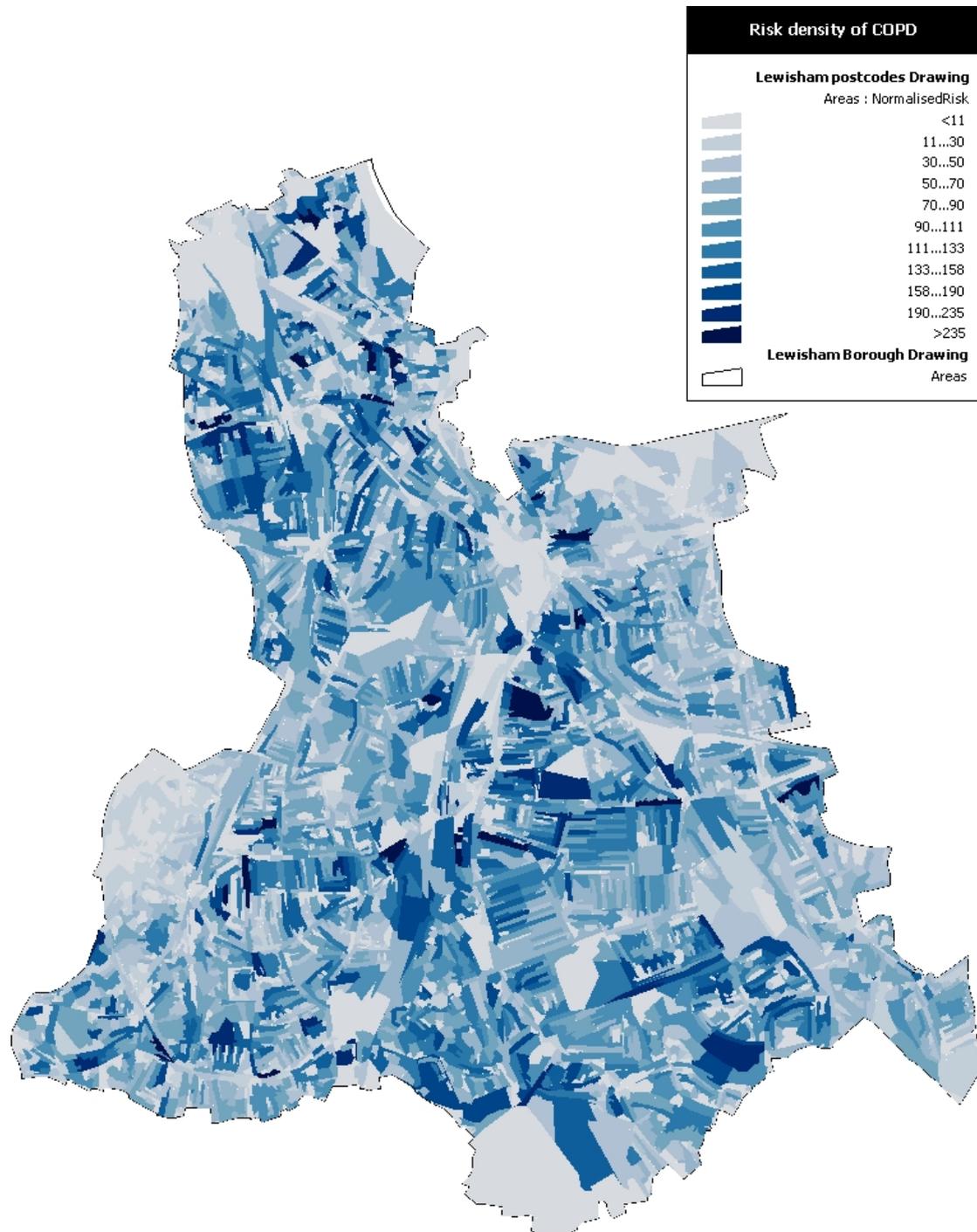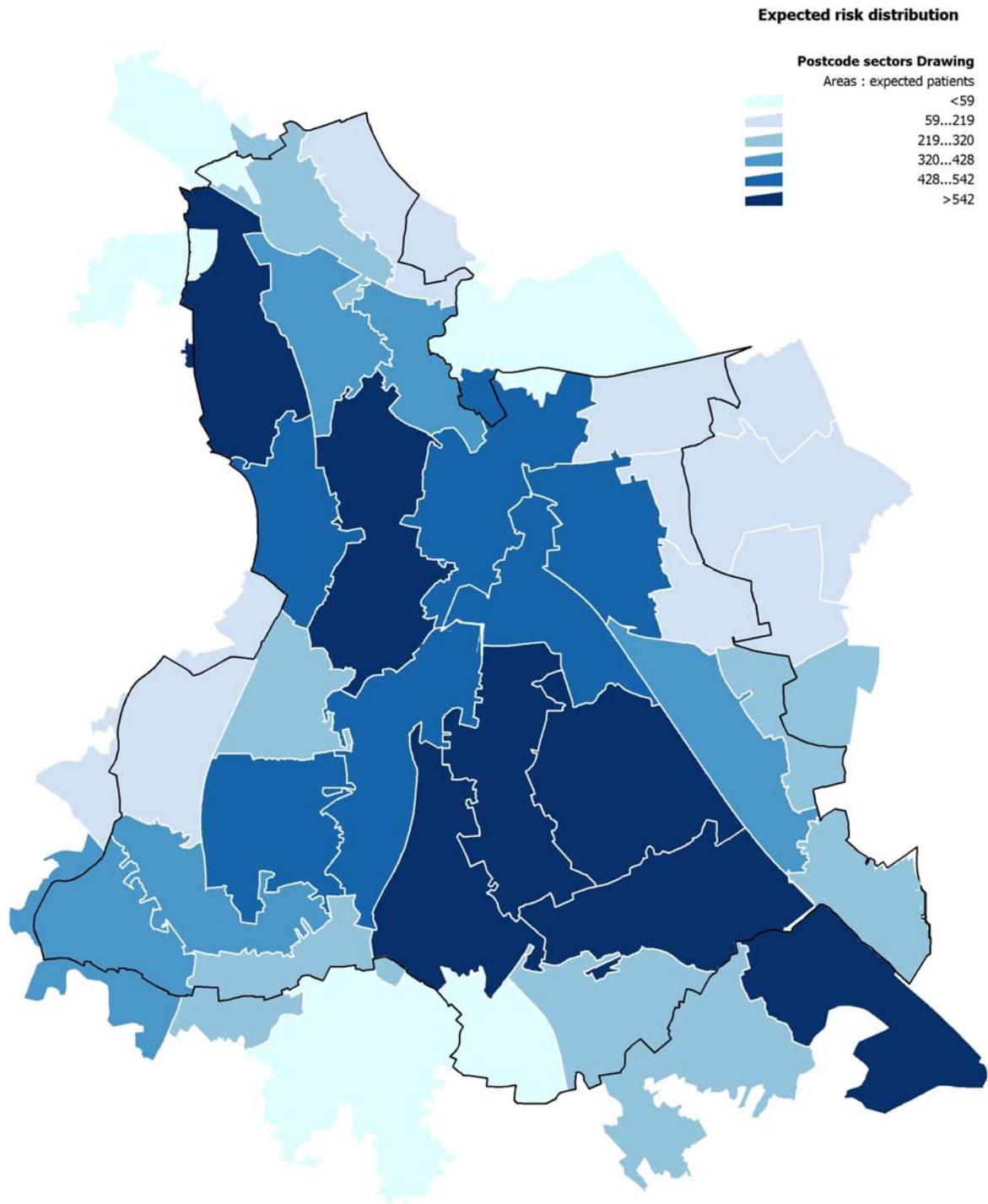**Figure 1**: **Risk density at postcode level**

Figure 2 shows the expected number of patients, mapped by postcode sector (eg SE8 12). Postcode sectors do not follow administrative boundaries, and the borough outline is shown in black. Where postcode sectors overlap the borough boundary the prevalence figures relate only to those parts of the postcode sector within Lewisham.

Figure 2: Expected COPD Prevalence by Postcode Sector

Expected risk distribution

Postcode sectors Drawing
Areas : expected patients
<59
59...219
219...320
320...428
428...542
>542

The pattern of highest expected prevalence is striking: a swathe runs from the north west of the borough to the south east, with a density gradient tending to go from the edges of the borough to its centre.

What is already known on this topic

Prevalence-modelled results for COPD are published annually at Primary Care Trust and GP Practice level[10], along with recorded prevalence from GP practice systems.

Underdiagnosis of COPD is apparently general, and is most pronounced in urban areas[Error! Bookmark not defined.], particularly in London.

A number of market research companies have developed health-related population segmentation and geodemographic tools, some of which purport to shed light on the detailed likelihood of particular health problems at very low levels of geography, even to individual postcode level. These tools depend on data which is proprietary, derived from market research surveys and retail spending patterns as well as publically available data such as population estimates by age and sex, ethnicity and socio-economic classification. The statistical analysis and the data on which it operates are not open to examination and verification, and have to be taken on trust.

What this study adds

The underlying model predicts individual numbers of cases in an area, a method that becomes less reliable as the size of the population decreases. It would be inappropriate to use it on the population of as small as a postcode. By aggregating the individual risks within an area and expressing the result as the relative probability of finding a case in the area, rather than predicting actual numbers of cases, case finding strategies based on looking where the risks are highest can be formulated.

The models described here are open to investigation and question, and the results can be aggregated to any defined geographical area. Output at postcode level enables GPs to identify where in their catchment areas they have the greatest likelihood of finding a previously undiagnosed patient with COPD, especially where their catchments have highly varied COPD risk densities.

Limitations of this study

Modelling provides estimates, not answers. Any model is strictly "wrong", but may be useful within certain limits. The methodology by means of which the models were

created does not easily permit the calculation of confidence intervals on the expected numbers of patients even though 95% confidence intervals were presented for the individual factors' odds ratios[11], so the models' outputs do not incorporate an indication of the strength of belief we can have in their results. . The risk factors were derived from data that may not be current. Ethnicity assumptions derived from census data almost ten years old may well be questionable, in the light of changing patterns of immigration in the period. Measurements of smoking prevalence suggest there has been a considerable drop in the proportion of people who smoke since the data on which the model is based were collected. As the smoking prevalence figures available are borough-wide, such a reduction would only affect the magnitude of the numbers, not the distribution across the borough. The smoking categories used, current smoker / former smoker / never smoked, take no account of any potential dosage effects.

While it would be possible to generate lists of individual patients with calculated odds ratios above a certain level, and use these to identify individuals for case finding, this risks falling foul of the ecological fallacy – that it is invalid to assume any individual bears the characteristics of the population of which (s)he is a member - and also may lead to problems of data protection and information governance. It is preferable to work at defined geographical levels.

Conclusions

The methods described offer a potentially usefully way forward for identifying undiagnosed cases. They are applicable to other long term conditions for which similar models are available on the APHO website. Ultimately, the validity or otherwise of the approach will depend on empirical results of its use.

Acknowledgements

References

[1] Nacul L, Soljak M, Meade T. Model for estimating the prevalence of chronic obstructive pulmonary disease: cross sectional data from the Health Survey for England. Popul Health Metrics 2007;5(1):8.

[2] NHS Lewisham, Preventing premature deaths in Lewisham: the contribution of primary care, Annual Report of the Director of Public Health 2009

[3] Nacul L, Soljak M, Samarasundera E, Hopkinson NS, Lacerda E, Indulkar T, Flowers J. Walford H, Majeed A, COPD in England: a comparison of expected, model-based prevalence and observed prevalence from general practice data, J. Public Health 2011, Vol33, No.1, 108-116

[4] Primatesta P, Stamatakis E, Health Survey for England 2001: Respiratory symptoms, atopic conditions and lung functions. London: The Stationery Office 2003

[5] http://www.statistics.gov.uk/statbase/Product.asp?vlnk=15381

[6] http://www.datadictionary.nhs.uk/web_site_content/supporting_information/nhs_postcode_directory.asp?shownav =1

[7] http://www.communities.gov.uk/publications/communities/indiciesdeprivation07

[8] www.manifold.net

[9] Ordnance Survey, ©Crown copyright and database right (2011)

[10] https://www.nhscomparators.nhs.uk

[11] Walford H, personal communication