PERGAMON

Social Science & Medicine 54 (2002) 931–937

SOCIAL SCIENCE & MEDICINE

www.elsevier.com/locate/socscimed

Predicting small area health-related behaviour: a comparison of multilevel synthetic estimation and local survey data

Liz Twigg*, Graham Moon

School of Social & Historical Studies, Institute for the Geography of Health, University of Portsmouth, Milldam, Burnaby Road, Portsmouth PO1 3AS, UK

Abstract

A recent paper in Social Science and Medicine (Twigg et al. 50 (2000) 1109) outlined an approach to the estimation of prevalences of small-area health-related behaviour using multilevel models. This paper compares results from the application of the multilevel approach with those derived using the more traditional strategy of the local 'lifestyle' survey. Estimations of smoking prevalence and high alcohol consumption are examined and critical assessments made of both estimation approaches. It is concluded that the alternative method is more suited towards the prediction of smoking rates as opposed to unsafe alcohol consumption.

Keywords: Smoking; Drinking; Synthetic estimation; Multilevel models; Health survey for England

The central importance of health related behaviour, such as smoking, diet, drinking and exercise, with regard to health outcome is now widely recognised. A number of key national and international policy documents have set out targets for either an increase or decrease in specific behaviours (WHO, 1981; DoH, 1992). Responsibility, for facilitating change or for providing counselling services is usually left to organisations or agencies working at the local area or neighbourhood level (e.g. community alcohol teams and primary care health promotion clinics). These local organisations and agencies have a need for locally specific data on health-related behaviours to assist in the targeting of their services and to provide a yardstick against which to monitor any changes resulting from their actions. Such information however, is not routinely available at the local scale. Consequently, alternative strategies for the estimation of small area health-related behaviour have evolved.

The main focus of this paper is to compare and contrast the results generated by two such strategies. We consider the most commonly used strategy — the local health survey — and a more recently devised approach based on multilevel models of national survey data (Twigg, Moon, & Jones, 2000). The paper is organised into three sections. The first section outlines the two strategies under discussion. A comparison of the estimates generated by the two methods is presented in the second section of the paper. We conclude with a brief discussion of the implications of the comparison in the final part of the paper. The two key themes that the paper pursues are the pervasive nature of uncertainty and the importance of geographical context in making estimates of small area health related behaviour.

Background

Although large scale, well-designed, annually repeated, continuous surveys such as the *General Household Survey* (GHS) and the *Health Survey for England* (HSE), provide information on some of the more important health-related behaviours such as smoking, diet, exercise and alcohol consumption, they do not provide reliable estimates below the standard region or health authority level, respectively (Colhoun & Prescott Clarke, 1996; ONS, 1996). One of the most basic strategies for obtaining estimates for local geographies

*Corresponding author. Tel.: 023-9284-2466; fax: 023-9284-2512.

E-mail address: liz.twigg@port.ac.uk (L. Twigg).

0277-9536/02/$ - see front matter
PII: S0277-9536(01)00065-X

has been to assume that these reported regional rates also apply at the local level. This approach does not however facilitate the identification of local variation and runs the risk of committing the ecological fallacy (Robinson, 1950) in that it denies the inevitable existence of variation within the region or health authority. Although the evidence concerning such variation is mixed (Duncan, Jones, & Moon, 1993, 1996, 1999), most would agree that it certainly occurs within the crude spatial units used in official surveys. To this end the shortcomings of this most basic of strategies are widely recognised.

The most common approach to generating useful local information is for local agencies or authorities to undertake their own survey of health-related behaviour. Several inadequacies can be noted with this method. These centrally concern the interaction of the resourcing of the survey and its design. An average-sized health authority might have expected in the mid-1990s to allocate at least £50,000 to undertake a single survey using a postal-based data collection approach and a well-presented multi-subject survey schedule administered by a reputable survey agency.[1] Attempts at cost-saving may result in a local survey that is not as robustly designed and delivered as its national counterpart; it may, for example, have non-standard question wording and, hence, a limited level of comparability with national data. More importantly however, even well-resourced local surveys seldom achieve representativeness at the sub-health authority level. Instead their sample sizes constrain their application to the generation of estimates of health-related behaviour for the whole health authority. The sample size implications of this shortcoming are considerable as a representative sample size for a whole health authority will be the same as that required for each of the component local areas within that health authority. These difficulties are compounded by the need for surveys to be repeated regularly to be of real use in monitoring change and assessing current needs.

Against these shortcomings must be placed the fact that local surveys provide a high profile manifestation of health promotion and public health activity within an area. They thus have uses for health authorities that may extend beyond the primary purpose of providing local data on health-related behaviour. They also provide the only means by which, however inefficiently or ineffectively, a health authority can collect data that relate directly to its population. These advantages to surveys both draw heavily on the notion of 'context' (Macintyre, MacIver, & Sooman, 1993). They indicate that place matters for health authorities insofar as the survey demonstrates the authority's concern for its area and assembles data that are specific to that area. Less advantageously, they generate a local contextuality as opposed to a national one; coverage of the country is far from complete and extant surveys remain largely in the domain of their sponsors and thus defy linkage and the generation of a more comprehensive comparative picture of health-related behaviour. The alternative possibility of collecting data as a by-product of the routine surveillance required as part of the national general practitioner contract is annulled by the diversity of data recording systems in general practice, the difficulties of data-sharing between practices and restrictions in the surveillance requirement.

A further alternative to the local survey as a strategy for estimating small-area health-related behaviour is to employ techniques of synthetic estimation. Synthetic estimation, under a range of names, has a considerable pedigree.[2] In essence the term refers to a family of techniques that derives the correlates of a particular behaviour or phenomenon from a large-scale survey or pre-existing study and applies those correlates to local data. For example, smoking behaviour in a large national survey might be cross-tabulated against gender, social class and marital status, with the proportions of smokers being derived for each sub-group. These proportions could then be applied to the numbers found in each sub-group in local areas (using the population census) to provide estimates of local smoking prevalence. A more refined method might use coefficients from generalised linear models rather than proportions from cross-tabulations.

In this paper we focus on synthetic estimation based on multilevel modelling.[3] The multilevel approach represents an advance on work of the form outlined in the previous paragraph in that it recognises that health-related behaviour simultaneously reflects both individual compositional factors and contextual, place-specific matters. Non-multilevel approaches to synthetic estimation have tended simply to apply national equations to local data; they have assumed that the national relationship between the behaviour in question and, typically, age, sex and social status, will be replicated at the local level. This assumption confounds national and local contextuality and assumes that the national context can

[1] Mean cost of district health authority 'lifestyle' surveys as reported in a telephone survey conducted by the authors with authorities known to have undertaken such a survey between 1995 and 1997.

[2] Approaches include iterative proportional fitting, entropy maximizing, regression interpolation and many others. For a useful review see Maxwell (2000). Examples of synthetic estimation which link survey and census data in the health field are given by Benzeval and Judge (1994) and Charlton (1998).

[3] See Goldstein (1995) or Kreft and de Leeuw (1998) for general studies of multilevel modeling. Jones (1991) provides an introduction for geographers.

represent the local. Multilevel approaches, in contrast, seek to control for the hierarchy of factors impacting upon behaviour and, crucially, they do so *simultaneously*, working on a single dataset structured to identify the individual, local, areal and regional 'levels' of influence. Furthermore, they can take into account the fact that there may be an interaction between levels. Thus, while a person may be more likely to smoke if she/he is from a lower social class, if that individual also resides in an area where there is a high percentage of working class people then the likelihood of smoking is even greater.

Additionally, a multilevel approach to synthetic estimation offers methodological advantages. The national datasets that provide the basis for the approach are typically based on hierarchical sampling structures in which selected individuals or households are chosen from higher level strata. Effective modelling requires recognition of multilevel processes and the autocorrelation consequent upon such hierarchical sampling designs. Owing to the use of precision-weighted estimation in model fitting, the multilevel approach is also relatively robust to variations in the number of observations in each sampling unit; model estimates based on relatively few observations are weighted towards the global average for the data.

The multilevel estimates of health-related behaviour examined in this paper derive from a two-stage approach to the estimation of smoking and unsafe drinking prevalence. There is a comprehensive report of this approach in Twigg et al. (2000) and the method will only be summarised here. First, multilevel equations for the two chosen outcomes were derived from an analysis of the 1994 Health Survey for England (Colhoun & Prescott Clarke, 1996). Individual predictor variables were age, gender and marital status while local-level contextual variables included tenure, class, and car-ownership. These choices reflected three imperatives: factors known to influence the behaviours in question, presence in the HSE and, in order to facilitate local estimation, presence in the population census. Within-level and cross-level interactions between predictor variables were assessed for significance and adjustment factors for each health authority in England were identified. Smoking prevalence was measured as whether or not a person currently smoked cigarettes. The UK Departments of Health have published guidelines relating to 'safe' levels of alcohol consumption and these currently stand at 3 to 4 units a day for men and 2 to 3 for women (Interdepartmental Working Group, 1995). Cut-off points of 14 units for women and 21 units for men were used for the work described here. The second stage of estimation involved 'reverse-engineering' the multilevel equations using small area census data in order to generate estimates of the behaviour at the local government ward level. This spatial level was chosen as it is also used in the routine reporting of mortality data. Providing estimates on health-related behaviours for the ward therefore facilitates efficient and straightforward data linkage for other aspects of health authority work; it is also important in the definition of health authority sub-divisions such as primary care group areas.

Despite their ability to take simultaneous account of individual and higher level factors, multilevel-based estimations of small-area health-related behaviour are, like local surveys, not without problems. In the case of the multilevel approach, these essentially derive from the methodological compromises necessary in working initially with a national survey with a fixed hierarchical structure, clear confidentiality constraints regarding disclosure and a particular content, and subsequently with a national census affected by the same issues but in different ways. To this end, the method of Twigg et al. (2000) was not only constrained to work with explanatory variables present with identical definitions in both the HSE and the population census, it also had to rely on the approximate spatial equivalence of local government wards to postcode sectors and, as a consequence of confidentiality, form local level indicators from data on surveyed individuals rather than the population data of the undisclosed local sampling units.[4] Additionally, a reduced set of individual-level predictors was necessitated by the limited number of detailed census cross-tabulations allowing enumeration of individual characteristics. Against these undoubted problems must be placed the significant cost advantage of an approach which uses routine data and can be replicated regularly using local demographic data and annual releases of the HSE. It is also an approach which provides national coverage.

To summarise, both surveys and multilevel synthetic estimators have significant problems as approaches to the generation of small-area estimates of health-related behaviour. Yet both also have advantages. Surveys can provide genuine local data but design and sampling deficiencies mean that they are seldom in reality the 'ground-truth' that other approaches should be expected to match. Multilevel synthetic estimators, in contrast, represent a cost-effective option but one that is undeniably technocratic and characterised by methodological compromise. Both approaches should be seen as being amongst a number of competing estimation strategies; neither need necessarily be 'correct'. In the following section, we compare and constrast results using the two strategies in order to exemplify empirically the issues raised above.

[4] The formation of higher level explanatory variables from sample data at level-1 also incurs penalties in terms of the standard errors of the eventual estimates. This impact is however, as yet, imperfectly understood.

Comparison

Obtaining local survey data to compare with multilevel synthetic estimates is itself a problematic task. Health authorities that have undertaken surveys must be identified and ward-level data obtained. Original survey instruments and sampling strategies and outcomes must be examined to ensure comparability of question wording and likely representativeness. Ideally, surveys selected for comparative purposes should be contemporaneous to the sources used to generate the synthetic estimates of health-related behaviour. Taking into account these difficulties, access to three surveys was negotiated: the Health of the Welsh (1996), the Newcastle and South Tyne Health Survey (1994) and Health Quest Portsmouth and South-East Hampshire (1993). These were all multi-purpose health surveys designed to obtain information on a range of health-related behaviours and other factors associated with personal health and the use of the National Health Service. Two of the comparator surveys were undertaken by external agencies on behalf of health authorities (Newcastle/South Tyne and Portsmouth and South East Hampshire). The Welsh survey was managed by Health Promotion Wales.

The wording used in the smoking questions was very similar across all three surveys. However the Newcastle and South Tyne data exhibited major differences in its definition of problem drinking and, in consequence, was only useful for comparison with smoking estimates. Several other known differences between the surveys were recorded. The Welsh data referred only to people aged 16–64 and had a known poor response to questions on problem drinking. The Portsmouth and South East Hampshire data, in contrast to that available from the other surveys, was not collected with ward-level sample representativeness in mind. Instead it was designed to be representative at a 'neighbourhood' level in which neighbourhoods were an amalgamation of wards. However this fact did not, in practice, prevent the reporting and use of survey results at the ward level by local health and social care workers.

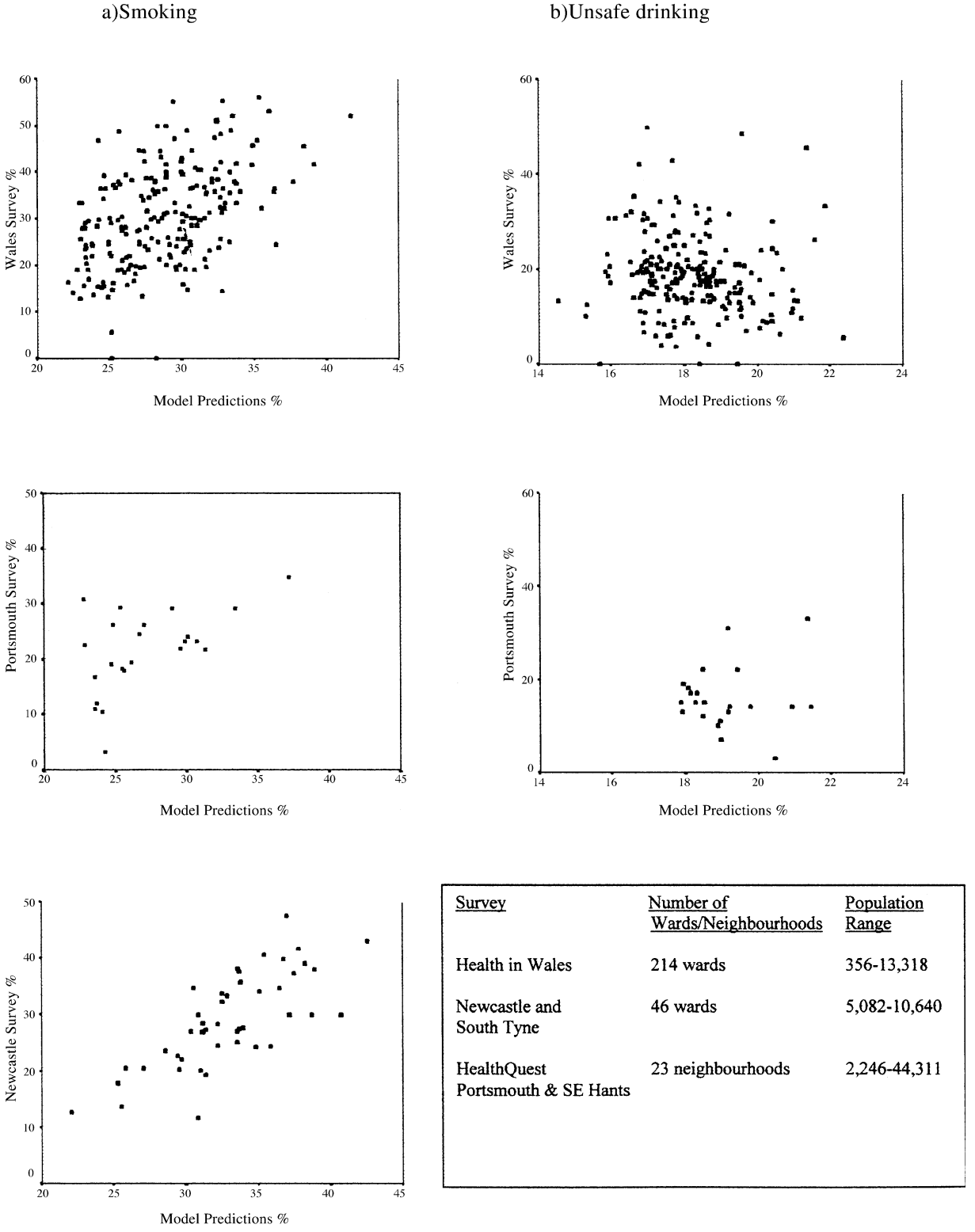
In all surveys, information was collected via a self-completion postal-returned questionnaire. The Welsh survey included over 15,000 respondents derived via a stratified multi-stage clustered design. The Portsmouth data were collected from 5100 randomly selected respondents via the age-sex register held by the relevant Family Health Services Authority (FHSA), as was the Newcastle and South Tyne sample of 9800. Overall response rates were 62% for Portsmouth, 67% for the Welsh survey and 69% for Newcastle and South Tyne.

Figs. 1a and b show the results of comparing the survey data with that obtained from multilevel synthetic estimation for, respectively, smoking and problem drinking. The multilevel parameter estimates of the two models that underpin the smoking and drinking predictions have been reported and discussed by Twigg et al. (2000). The table in Fig. 1 also lists the number and population size of the local government wards or neighbourhoods used in each of the surveys. Across England the average adult population of a local government ward in 1991 was approximately 4300.

Both sets of figures (i.e. smoking and drinking) employ similar axial scales in order to accentuate departures from congruity. It is clear that there are substantial disparities between the two estimation methods and between the two behaviours. Differences between survey estimates and multilevel synthetic estimates are least marked in the case of smoking. Indeed, in purely correlational terms, the fit between the two sets of data on smoking is strong in all three comparator settings (Wales $r = 0.491$, Portsmouth $r = 0.549$, Newcastle $r = 0.749$; all significant $p > 0.05$). Examination of Fig. 1a reveals however that this strong relationship hides a number of interesting facets. First, the multilevel synthetic estimates suggest a range of smoking prevalences that is far smaller than that suggested by surveys. This is to be expected given that any local synthetic estimation procedure is based on information from a wider population. Second, and as a consequence of the variation in estimates of the range of smoking prevalences, the multilevel estimates tend not to identify the places that surveys deem to have the highest and the lowest prevalences. In contrast, it seems that places with low prevalences according to surveys have prevalences inflated by some 20% using the multilevel estimates while places with high survey prevalences are deflated by around 10%. Third, and more reassuringly, there is a relatively good match between the estimates from the two approaches for places which appear to have levels of smoking around the national mean of ca 28%. Finally, with the possible exception of some of the extreme values in the Welsh case, there is little evidence that any individual wards have prevalence estimates from the two approaches that are actually contradictory. Whilst the synthetic estimation approach did not generate any zero values, Welsh survey results give zero prevalences for smoking. These are entirely expected and an important aspect of survey response. The zero prevalences are, in fact, reported for rural wards where response rates were particularly low for these questions. They reflect total smoking abstinence amongst the small numbers of individuals who responded in those wards. If the two zero smoking prevalence wards are removed from the dataset, the resultant correlation coefficient ($r = 0.494$) is similar to that reported above.

The situation with the problem drinking estimates is indicative of far more substantial differences between the two approaches to prevalence estimation. Even in correlational terms, there is a large disparity between the survey based predictions and those derived from the

Figure 1



Survey	Number of Wards/Neighbourhoods	Population Range
Health in Wales	214 wards	356-13,318
Newcastle and South Tyne	46 wards	5,082-10,640
HealthQuest Portsmouth & SE Hants	23 neighbourhoods	2,246-44,311

Fig. 1. Survey results versus model predictions.

multilevel synthetic estimators. For the two settings where comparison was possible, correlations failed to reach statistical significance and thus indicate no relationship between the two estimations derived from the two approaches (Portsmouth $r = 0.072$; Wales $r = -0.128$). The negative coefficient for the Welsh data suggests that what correlation is present tends to indicate that higher values on one approach match lower ones on the other. Inspection of the scatterplots (Fig. 1b) suggests that surveys indicate a far greater range of problem drinking prevalences. There are places where the multilevel synthetic estimation approach suggests prevalences of between 15 and 20% but surveys suggest no problem drinking. Conversely areas with high problem drinking according to surveys are estimated to have rates 10% lower on the multilevel estimates for Portsmouth and up to 30% lower on the multilevel estimates for Wales. Again, zero prevalences of problematic alcohol consumption are reported for three wards in the Welsh survey results and again these relate to wards where response rates were particularly low. Removal of these from the dataset results in a correlation coefficient of -0.140. Whilst this value is similar to that reported above, the removal of the zero prevalences results in the correlation reaching significance ($p > 0.05$). Yet the correlation remains negative and the problematic contradiction between the survey and multilevel estimates persists.

Discussion and conclusions

Clearly the disparities between the estimates based on surveys and on multilevel synthetic estimation are of concern. A starting point in this paper has been the view that neither one approach nor the other necessarily represents a 'gold standard'. We cannot be certain that either is 'right'. As a simpler, less technocratic and more established approach, it may be that the survey estimates are seen as being more believable but this conclusion is not necessarily sound. In this final section we consider the implications of the findings set out in the previous section for both approaches.

Taking the survey estimates first, all three comparator surveys were designed to collect samples that would be statistically representative at the sub-health authority level. While this, of course, represents the case for surveys and against an alternative approach to obtaining local level information on health related behaviour, it also indicates a limitation to survey data as a sound basis for comparison in situations where statistical representativeness is not claimed for the survey. Though the results for the Portsmouth and South-East Hampshire survey were adequate for comparison, representativeness was not achieved at the ward level. Consequently, further linkage to other routine data was not easily possible. Further, and notwithstanding their representativeness in terms of design, surveys also experience problems of under-response. With the Welsh data for example, the response rate for the problem drinking variable was one of the worst in the whole survey (pers. comm.). In part, this under-response may be indicative of a general problem: to some extent in contrast to smoking, it is well-documented that collecting information on alcohol consumption expressed as units is problematic and that self-assessed levels of consumption are particularly prone to measurement and bias error.

With regard to the multilevel synthetic estimates, limitations in the modelling process have already been noted. One further point can be made however. The evident bunching of the multilevel synthetic estimates around the overall mean for the particular behaviour is, to some extent, an inevitable consequence of the modelling process. First, a significant proportion of the highest and lowest prevalence rates from surveys reflect a denominator problem whereby extreme rates stem in part from low base numbers. In multilevel modelling these data points are shrunk towards the global mean of the dataset. This 'borrowing strength' has the effect of forcing multilevel synthetic estimates into possibly artificially narrow bandings. Second, when using the technique of multilevel modelling, the user is able to assess the variation that remains unexplained at various levels. In the alcohol model of Twigg et al. (2000), there was a much larger residual ecology present at higher geographic levels. This could be due to missing explanatory variables or to the inability of the model-based estimation process to take account of the complex ecological and cross-level variation that is known to characterise problem drinking (Duncan, Jones, & Moon, 1998). Its effect on the estimation process was to constrain further the range of values that could be taken by the problem drinking indicator. The smoking indicator was less affected as smoking patterns are less idiosyncratic and more amenable to summary within the constraints imposed by the synthetic estimation modelling process.

A final, separate, point to consider concerns the comparison between the multilevel synthetic estimates and the Welsh survey data. In Wales, census data were unavailable for a small number of wards due to their sparse populations and synthetic estimates could not therefore be generated for these areas. Moreover, whereas for English wards it was possible to introduce an adjustment factor into the estimation process using the residuals for the health authority to take account of wider contextual effects, the Welsh health authorities were not sampled in the HSE and no adjustment could therefore be made in the final predictions. Of course, it should also additionally be noted that an interesting theoretical question is raised by the construction of

predictions for Wales using equations derived from the multilevel modelling of English (HSE) data. It remains a possibility that any lack of correlation between multilevel synthetic estimations and Welsh survey data simply reflects national differences in the social correlates of drinking and smoking.

To conclude, the results of this work confirm that local surveys and multilevel synthetic estimates have both advantages and disadvantages as strategies for generating local estimates of smoking and problem drinking. It is certainly possible to use multilevel modelling to generate estimates for far less cost than a local survey but the modelling process requires significant compromises with the complexity of health-related behaviour. If congruence between local survey data and synthetic multilevel estimates is seen as a desirable goal, it is clear that multilevel synthetic estimation works very much better for smoking than for problem drinking and is, perhaps, better suited to ordering wards in terms of prevalence rather than estimating precise levels of prevalence. A local area survey, that is designed to be reliable at ward level, uses a valid field instrument and is characterised by high response rates can be regarded as the 'gold standard'. However, the presence of all three attributes in any one local survey is very rare. If this standard is unavailable then a local health planner may be better advised to use the multilevel synthetic estimation approach rather than the 'random noise' generated by poorly designed, poorly answered local surveys.

Acknowledgements

The authors acknowledge the support of the ESRC and the assistance of Martin White, Simon Raybould, Paul Iggulden, Chris Roberts and Laurence Moore.

References

Benzeval, M., & Judge, K. (1994). The determination of hospital utilisation: Implications for resource allocation in England. *Health Economics*, *3*, 105–116.

Charlton, J. (1998). Use of the census Samples of Anonymised Records (SARs) and survey data in combination to obtain estimates at local authority level. *Environment and Planning A*, *30*, 775–784.

Colhoun, H., & Prescott Clarke, P. (1996). Survey methodology and documentation, Health Survey for England 1994 (Vol. 2). London: HMSO.

Department of Health. (1992). *The health of the nation: A strategy for health in England*. London: HMSO.

Duncan, C., Jones, K., & Moon, G. (1993). Do places matter? A multilevel analysis of regional variations in health-related behaviour in Britain. *Social Science and Medicine*, *37*, 725–733.

Duncan, C., Jones, K., & Moon, G. (1996). Health-related behaviour in context: A multilevel modeling approach. *Social Science and Medicine*, *42*, 817–830.

Duncan, C., Jones, K., & Moon, G. (1998). Context, composition and heterogeneity: Using multilevel models in health research. *Social Science and Medicine*, *46*, 97–118.

Duncan, C., Jones, K., & Moon, G. (1999). Smoking and deprivation: Are there neighbourhood effects? *Social Science and Medicine*, *48*, 497–505.

Goldstein, H. (1995). *Multilevel statistical models*. London: Edward Arnold.

Interdepartmental Working Group. (1995). *Sensible drinking*. Department of Health, London.

Jones, K. (1991). *Multi-level models for geographical research. Concepts and techniques in modern geography* No. 54. Norwich Environmental Publication, University of East Anglia.

Kreft, I., & de Leeuw, J. (1998). *Introducing multilevel modelling*. London: Sage.

Macintyre, S., MacIver, S., & Sooman, A. (1993). Area, class and health: Should we be focusing on places or people? *Journal of Social Policy*, *22*, 213–234.

Maxwell, J. C. (2000). Methods for estimating the number of "hard core" drug users. *Substance Use and Misuse*, *35*, 399–420.

ONS. (1996). *General Household Survey*, 1994 (*computer file and user guide*). Data Archive, Colchester.

Robinson, W. S. (1950). Ecological correlations and the behaviour of individuals. *American Sociological Review*, *15*, 351–357.

Twigg, L., Moon, G., & Jones, K. (2000). Predicting small-area health-related behaviour: A comparison of smoking and drinking indicators. *Social Science and Medicine*, *50*, 1109–1120.

WHO. (1981). Global strategy for health for all by the year 2000. In: *Health for All Series*, No. 3. Geneva: WHO.