Impact case study (REF3b)

Institution: Plymouth University

| Unit of Assessment: Mathematical Sciences B10 |

| Title of case study: Bayesian methods for large scale small area estimation (SAE). |

1.   Summary of the impact (indicative maximum 100 words)
Small area estimation (SAE) describes the use of Bayesian modelling of survey and administrative data in order to provide estimates of survey responses at a much finer level than is possible from the survey alone. Over the recent past, academic publications have mostly targeted the development of the methodology for SAE using small-scale examples. Only predictions on the basis of realistically sized samples have the potential to impact on governance and our contribution is to fill a niche by delivering such SAEs on a national scale through the use of a scaling method. The impact case study concerns the use of these small area predictions to develop disease-level predictions for some 8,000 GPs in England and so to produce a funding formula for use in primary care that has informed the allocation of billions of pounds of NHS money. The value of the model has been recognised in NHS guidelines. The methodology has begun to have impact in other areas, including the BIS 'Skills for Life' survey.

2.   Underpinning research (indicative maximum 500 words)
The Statistics Group at Plymouth specialises in Bayesian modelling methods, with expertise in applying the necessary computationally intensive methods, using modern High Performance Computing (HPC) where necessary. Our work on survey methods, spatial statistics, and disclosure control methods gives us a strong understanding of the sources and structures of data that underpin our SAE work. Our applied modelling interests have led to the development of a methodology for a range of response types, including new frameworks for modelling multinomial response and multivariate response types by means of latent structure models, quantile regression methods, and spatial methods. This enables a wide range of applications in such areas as disease mapping and financial analysis. One early example of such modelling work was applied to cancer mapping [Ref.1]. The underlying methodology has continuously been developed at Plymouth within the Statistics group and has been applied to other problems.   Other established work by statisticians at Plymouth concerns disclosure control. Administrative data are only released with limitations on the ability to deduce details about individuals. Hence, small area predictions require the reconstruction of the multiway tables underlying a number of two and three way tables that have been released under the limits for disclosure control. Our works in this area (particularly [4]) underpin this reconstruction and therefore are key for the estimation.

Our SAE work is exceptional and distinctive in that:

- We fit relatively complex models to survey data (which may consist of several thousand person types), and can thus obtain a more sensitive understanding of the risk surface [1],

- Our understanding of disclosure control methods enables us to derive usable information on these person types based on released administrative data [2,3,4],

- We simulate predictive distributions for the risk of disease for these person types in 6,781 middle-level super-output areas in the whole of England,

- And we then re-apportion such estimates to some 8,000 GP practices. [5,6]

This is a computationally demanding task. There is a growing literature on SAE but, without our ability to apply it to the whole of England, these methods would not be trusted to inform decisions on a national scale. In SAE, our work out-performs that of others. The sheer volume of calculations requires the use of High Performance Computing to throughput all the calculations. Given fine-grained predictions, we can re-aggregate to whatever geography is of interest, which currently enables us to study access to healthcare.   In order to produce predictive distributions for small areas, the necessary data are not easily available for data protection reasons. For example, we require not only posterior distributions from survey models but auxiliary data (individual level data) on the 6,781 middle-level census output areas for the varied demographic information that

matches the census and other administrative data with the survey data used in the modelling. For disclosure control reasons, such data are not published. The research leading to impact here involves the reconstruction of multiway tables for a number of two- and three-way tables that have been released (with adjustments to limit disclosure control). The underpinning work in this area (see [4]) therefore is a key part of the estimation process, enabling us to quantify uncertainty in the auxiliary variables.

3. References to the research (indicative maximum of six references)

[1] **Hewson P.J.** and T.C. Bailey (2010): Modelling multivariate disease rates with a latent structure mixture model, *Statistical Modelling 10(3): 241-164.*

[2] **Burridge, J**. (2003): Information preserving statistical obfuscation, *Statistics and Computing 13:321-327.*

[3] Franconi, L. and **Stander, J.** (2002): A model based method for disclosure limitation of business microdata, *Journal of the Royal Statistical Society, Series D 51: 51-61.*

[4] Polettini, S., Franconi, L. and **Stander, J.** (2002): Model based disclosure protection. In Domingo-Ferrer, J. (Ed.) *Inference Control in Statistical Databases: from Theory to Practice.* Berlin: Springer-Verlag, pp. 83-96. (Peer reviewed book chapter)

[5] Asthana, S., Gibson, A., Bailey, T., Dibben, C., **Hewson, P**., Economou, T., Batchelor, D., Eastham, J., Craig, R., Scholes, S., Flowers, J., Jenner, D. Person (2008): Based Resource Allocation (PBRA)*: The Feasibility of Developing a Need-Based Approach to PBRA*. Report to the Department of Health (Policy Research Programme). University of Plymouth. 118pp.

[6] S.Asthana, A. Gibson, **P. Hewson**, T. Bailey, C. Dibben (2011): General practitioner commissioning consortia and budgetary risk: evidence from the modelling of fair shares practice budgets for mental health', *Journal of Health Services Research & Policy 16:95-101.*

4. Details of the impact (indicative maximum 750 words)

The research carried out in the area of statistical disclosure control has led to significant changes in the dissemination of micro data in official statistics. The seminal paper by Franconi and Stander [3] and subsequent work in the area of model-based micro-data protection (see [4]) have successfully resulted in the release of micro-data for research purposes, with respect to the Italian sample of the *Structure of Earning* survey. This survey, leading to a linked employer-employee database harmonized at European level, is now going to reach its third wave of release. Economists have carried out several studies, using the micro data file for research distributed at national level by Istat, and at EU level by Eurostat.

The ideas contained in another paper, published by researchers from Plymouth University in the field of statistical disclosure control, Burridge (2003) [2], have opened the field to the use of perturbation methods showing sound statistical properties by construction. Again, a new micro data file for research purposes on the system of account will be released in the near future, stemming from the work carried out at Plymouth University. Moreover, the IPSO method (Information Preserving Statistical Obfuscation) has also been implemented in the mu-Argus software (mu-Argus, 2013), which has been designed to create files of individual data to be released either for research purposes or for the public (1). Many national statistical offices in Europe and statistical agencies around the world use this software. Some of the developments contained in Polettini and Stander (2005), leading to an alternative evaluation and a related accurate approximation of the risk of re-identification, have also been implemented in the mu-Argus software.

Given a good understanding of administrative data systems, we have been able to contribute to problem solving within the UK. For example, 13.9% of the total primary care budget is directed to mental health, a total of £8 billion. The indicative allocation of several billion pounds of public money to GP practices, to provide services for mental health, is informed by what the NHS refers to as the 'Plymouth Model'. In the user guide [page 10](DoH, 2009):

'For mental health, the toolkit includes an entirely new methodology developed by Plymouth

University specifically for practice based commissioning. This new approach moves away from modelling historic utilisation and estimates need directly based on different person types' (2).

In order to do this, we have required state-of-the-art epidemiological models applied to the Health Survey for England, working on case mix classifications as well as a statistical reconstruction of the census data, known to be subject to statistical disclosure control. Quoting the NHS (DoH, 2009):

'The new methodology has undergone extensive testing by the researchers and DoH and we believe it provides a step-change improvement in the way we model mental health need' (2).

This work has been reported to select committees in parliament and recognised by MPs, all of whom have noted the size of GP commissioning groups (3).   It has also been fully considered by the National Audit Office, influencing the way public sector funding may evolve in the future (4). Finally, our expertise with national-scale small-area estimates has had further impact on work carried out for the 2011 Skills for Life Survey, which has been made available by the Department for Business, Innovation and Skills (5).

5. Sources to corroborate the impact (indicative maximum of 10 references)

(1) Evidence of the impact of work on Disclosure control: Mu-Argus Software
http://neon.vb.cbs.nl/casc/..%5Ccasc%5Cmu.htm; the manual is available at
http://neon.vb.cbs.nl/casc/..%5Ccascprivate%5Cdeliverables%5CMUManual4.3.pdf),

(2) Evidence of the impact of the work on funding GPs for mental health provision:

Department of Health (2009) Practice Based Commissioning: budget guidance for 2009/10: Methodological changes and toolkit guide DH (see

http://www.dh.gov.uk/prod_consum_dh/groups/dh_digitalassets/documents/digitalasset/dh_094392.pdf

(3) Asthana, S., Gibson, A. (2010), Funding Implications for Rural PCTs of the new NHS Resource Allocation Methodology, in All Party Parliamentary Group on Rural Services, The implications of national funding formulae for rural health and education provision, Report, Written and Oral evidence. London, House of Commons.

(4) National Audit Office (2011) 'Cross-government landscape review Formula funding of local public services ' REPORT BY THE COMPTROLLER AND AUDITOR GENERAL

HC 1090 SESSION 2010–2012

(5) Small area estimation applied to skills for life

BIS Research Paper Number 81C (2012) 2011 Skills for Life Survey: Small Area Estimation Technical Report, Department of Business, Innovation and Skills