Institution: University of Bristol

Unit of Assessment: 10 – Mathematical sciences

Title of case study: The use of multilevel statistical modelling has led to improved evidence-based policy making in education and other sectors

1. Summary of the impact (indicative maximum 100 words)

Since 2008, statistical research at the University of Bristol has significantly influenced policies, practices and tools aimed at evaluating and promoting the quality of institutional and student learning in the education sector in the UK and internationally. These developments have also spread beyond the education sector and influence the inferential methods employed across government and other sectors. The underpinning research develops methodologies and a much-used suite of associated software packages that allows effective inference from complicated data structures, which are not well-modelled using traditional statistical techniques that assume homogeneity across observational units. The ability to analyse complicated data (such as pupil performance measures when measured alongside school, classroom, context and community factors) has resulted in a significant transformation of government and institutional policies and their practices in the UK, and recommendations in Organisation for Economic Co-operation and Development (OECD) policy documents. These techniques for transforming complex data into useful evidence are well-used across the UK civil service, with consequent policy shifts in areas such as higher education admissions and the REF2014 equality and diversity criteria.

2. Underpinning research (indicative maximum 500 words)

Multilevel statistical modelling, a sophisticated methodological approach to data analysis, has been developed at the University of Bristol since 2005. This statistical technique is crucial in analyses of complex data sets that have non-trivial geographical and/or temporal components, data structures which could cause dangerous confusion if analysed using more basic statistical techniques.

The impact of these approaches has been maximised through deployment of research such as [2] in new versions of the hugely popular MLwiN statistical software package, along with new software such as REALCOM-impute [3] which carries out multiple imputation of missing data in the context of a multilevel model, MLPowSim to perform sample size calculations for multilevel modelling, and Stat-JR, an innovative software environment for promoting interactive statistical modelling.

Recent research has focussed on developing inference methods for realistically complex multilevel models, and developing an understanding of the computational methodologies required to fit complex models to large data sets, such as those obtained from national level data collection of, for example, pupil test scores and university applications.

Novel statistical methodology underpinning the impact:

- Introduction of re-parameterisation techniques within Markov chain Monte Carlo (MCMC) algorithms for fitting multi-level models, allowing the statistical analyses to become sufficiently efficient to be carried out on significantly larger and more complex data structures [2].
- Simulation-based graphical approaches to communicating statistical uncertainty of group effects in multilevel models [6], which are otherwise difficult to interpret and communicate.
- Novel multilevel modelling formulations for complex non-hierarchical data structures [5]. This type of data set arises, for example, when effects of both schools and neighbourhoods are to be taken into account in models of pupil performance.
- Advanced multilevel modelling for analysis of longitudinal data with multilevel structure. In particular when the observational units change, but the grouping units do not, such as clinical outcomes in a hospital in successive time periods, or pupils in a school in successive years [6].
- Multilevel statistical techniques for modelling multivariate data with different response types at several levels, and handling correlated measurement and misclassification errors [4].

Novel computational implementations underpinning the impact:

- The general computational capabilities of MLwiN, and the MCMC module in particular [1], have been updated to enable efficient inference to be carried out with complex models and large data sets. Key parts of this update, which have been in versions of MLwiN since 2.13 (released in August 2009), are based on the techniques of [2].

- Production of three new software packages: REALCOM-impute [3], to allow imputation of missing data; MLPowSim, to perform sample size calculations within the multi-level modelling framework; and Stat-JR, a software environment for promoting interactive statistical modelling.

The team that carried out the research described above moved to Bristol in 2005 (Goldstein, Prof. of Social Statistics, Steele, Prof. of Social Statistics, and Rasbash, Prof. of Computational Statistics), 2007 (Browne, Prof. of Biostatistics) and 2009 (Leckie, Senior Lecturer in Social Statistics).

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

[1] Browne, W.J. (2012) MCMC Estimation in MLwiN, v2.26. Centre for Multilevel Modelling, University of Bristol. http://www.bristol.ac.uk/cmm/software/mlwin/download/2-26/mcmc-web.pdf
*[2] Browne, W.J., Steele F., Golalizadeh, M., and Green M.J. (2009) The use of simple reparameterizations to improve the efficiency of Markov chain Monte Carlo estimation for multilevel models with applications to discrete time survival models. *Journal of Royal Statistical Society*, A, **172**, 579-598. DOI:10.1111/j.1467-985X.2009.00586.x
[3] Goldstein, H. (2011) REALCOM-IMPUTE, User Guide. Centre for Multilevel Modelling, University of Bristol. http://www.bristol.ac.uk/cmm/software/realcom/imputation.pdf
[4] Goldstein, H., Kounali, D. and Robinson, A. (2008) Modelling measurement errors and category misclassifications in multilevel models. *Statistical Modelling*, **8**, 243-261. DOI:10.1177/1471082X0800800302
*[5] Leckie, G. (2009) The complexity of school and neighbourhood effects and movements of pupils on school differences in models of educational achievement. *Journal of the Royal Statistical Society*, A, **172**, 537-554. DOI:10.1111/j.1467-985X.2008.00577.x
*[6] Leckie, G. and Goldstein, H. (2009) The limitations of using school league tables to inform school choice. *Journal of the Royal Statistical Society*, A, **172**, 835-851. DOI:10.1111/j.1467-985X.2009.00597.x
* references that best indicate the quality of the underpinning research.

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

Since 1 Jan 2008, MLwiN has been purchased by 613 non-academic individual users and 75 organisations (67 site licenses for 50 users, and 8 for 250 users). It has also been downloaded for free by 3,846 UK academics, and purchased by 5,518 overseas academics. The Bristol Centre for Multilevel Modelling website is widely acknowledged as the premier resource for research and training in multilevel modelling, with around 1,100 page-loads and 360 unique visitors per day (65% of whom are from outside the UK). The on-line Learning Environment for Multilevel Modelling (LEMMA), launched in April 2008 as part of the Economic and Social Research Council National Centre for Research Methods, now has around 10,600 registered users, of which 70% are international and 14% are non-academic. The sheer number of users demonstrates the reach and significance of the underpinning research [1-3]. Non-academic users of Centre software, and hence the underlying research, include the World Health Organisation, Statistics Canada, Statistics Norway, Netherlands Central Bureau of Statistics, the UK Departments for Education, Health, and Work and Pensions, the Scottish Executive, the Office for National Statistics and the Higher Education Funding Council for England. Analyses by these bodies, and others, as well as by the academic users of the software, have significant impact on society; we focus on three specific areas.

A. **Impact on UK and international policy and public awareness relating to measuring educational effectiveness and school performance**

Improved school evaluation policies: In the UK, Goldstein's multilevel modelling framework (including, for example, [4-6]) has provided the statistical toolkit which has provided evidence to inform and influence key national policies related to school evaluation such as: the utility of school self-evaluation, national pupil databases, contextualised value-added (CVA) measures of school performance, and separate value-added measures for different student groups (introduced by the Department for Education (DfE) 2011). The research has also promoted the use and understanding of a wider range of outcomes and measures by DfE, the Department for Children,

School and Families, the Office for Standards in Education, Children's Services and Skills (Ofsted) and the Learning and Skills Council [a]. MLwiN is currently used within DfE to calculate published measures of CVA school performance, an integral part of the Ofsted school inspection process, and also to construct the Learning Achievement Tracker, a new tool for schools and further education colleges to appreciate progress made by students since the end of compulsory schooling [b]. In particular "MLwiN allows … complex cross-classified multilevel structures, and these, using, in the main, MCMC methods, were used to inform the DfE about the variations in pupil performance associated statistically with their social background and school attended," [b] which demonstrates that research items [1,2,5] are having significant impact in this area.

Public understanding of league tables: The statistical research carried out by [6] demonstrates the limitations of using the government's school league tables to inform school choice. This has promoted public understanding of the problems of league tables through widespread communication to non-academics via popular articles and other media, including interviews for the BBC Radio 4 programmes `Analysis' and `The Learning Curve' and articles in the Financial Times, Daily Telegraph and Times Educational Supplement [c]. This work demonstrates impact in terms of both reach and significance, given that it has been incorporated into policy documents by numerous governments and non-governmental organisations, both overseas (including the OECD) and UK (including the National Union of Teachers and the Institute for Government), to influence public thinking and new policy development on educational accountability and improvement issues [d].

Improved understanding of rural educational issues: A further example of the importance of strong statistical research [4] with associated software [3] for educational research resulting in societal benefit is a recent Department for Environment, Food and Rural Affairs report [e] in which the new sophisticated methods were used to investigate whether higher attainment of rural school pupils was symptomatic of a better educational environment or simply a by-product of generally higher social position in rural areas. Missing data, and multilevel structure, are endemic in such a study, and [3] was necessary since "the use of multiple imputation in this study should provide more accurate [analyses] than a complete case analysis [throwing out records with missing data], and should also increase the power of the analyses so that small differences between settlement types can be more easily distinguished" [e].

B. **Admissions to UK universities**

The Schwartz Report on Fair Admission (2004) was instigated to review "options which English institutions providing higher education should consider in assessing the merit of applicants for their courses." Consequently, the Supporting Professionalism in Admissions programme was set up to support higher education institutions to develop admissions policies ensuring fair access. The programme has recently carried out a major investigation into the use of so-called contextual data to inform admissions decisions, publishing a 2012 report [f]. The data supporting the review was multilevel in nature, and extensive (between 0.4 and 1.6 million individual school records). "With such a sensitive issue, it was important to fit a statistical model that would be robust to criticism" [g]. A recent version of MLwiN, which incorporates the latest MCMC module [1] incorporating the advanced MCMC techniques developed by [2] "was necessary to achieve satisfactory results with such a high volume of data and highly correlated variables of interest" (type of school and educational attainment being two key candidate explanatory variables for degree performance) [g]. The report concludes that the type of school is an important predictor of degree performance. The recommendation to HEIs is therefore to incorporate such contextual information in their admissions decisions, thus having an impact across the UK in terms of accessibility of higher education.

C. **Equality and diversity policy for REF2014 submissions**

Analysis of RAE2008 data was carried out by the Higher Education Funding Council for England using MLwiN. One aspect of this was a multilevel analysis over 30,000 records with a binary response to indicate whether an individual was included in the RAE2008 submission. "To carry out such a large multilevel analysis with binary response data required the use of the recent optimised MCMC components of MLwiN" [1], based on the research of [2] [g]. A key finding was of selection biases against certain ethnic groups that were not explained by controlling for other factors [h]. In response, significantly improved rules on equality and diversity have been introduced for REF2014 (http://www.ref.ac.uk/equality/), thus having a significant impact on UK Higher Education.

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

[a] Director General, Monitoring and Assessment, UK Statistics Authority has provided information about influence of University of Bristol research on government and public understanding of UK policy on school evaluation.

[b] Consultant Statistician to DfE.
May be contacted to corroborate the influences of Centre for Multilevel Modelling team's value-added studies on and use of MLwiN by DfE statisticians and Ofsted.

[c] Sources to corroborate engagement with public.
http://news.bbc.co.uk/1/shared/spl/hi/programmes/analysis/transcripts/30_01_12.pdf
http://www.bbc.co.uk/radio4/factual/learningcurve_20080616.shtml
http://www.ft.com/cms/s/0/17dfb862-7ad4-11de-8c34-00144feabdc0.html
http://www.telegraph.co.uk/education/secondaryeducation/6005906/Grammar-schools-penalised-by-new-league-tables.html
http://www.tes.co.uk/article.aspx?storycode=6009334

[d] Wildeman (2011) "Beware of the misleading means and measures". In *Transformation Audit 2011*, published by The Inclusive Economies Project, which is located within the Policy and Analysis Unit of the Institute for Justice and Reconciliation (IJR). http://transformationaudit.org/blog/wp-content/uploads/2012/02/Opinion-Beware-of-the-misleading-means-and-measures.pdf
Mulgan (2010) "Transparency Occasional Paper 1: Transparency and Public Sector Performance". Queensland Office of the Information Commissioner and the Australia and New Zealand School of Government working paper, http://www.anzsog.edu.au/media/upload/publication/93_1-Mulgan-Transparency-and-Public-Sector-Performance.pdf
Masters, Rowley, Ainley and Khoo (2008) "Reporting and comparing school performances". Commissioned by the Reporting and Accountability Branch, National Education Systems Group, Commonwealth Department of Education, Employment and Workplace Relations (DEEWR) http://apo.org.au/research/reporting-and-comparing-school-performances
OECD (2008) "Measuring improvements in learning outcomes: Best practices to assess the value-added of schools." Organisation for Economic Co-operation and Development, DOI:10.1787/9789264050259-en
Rosenkvist, M.A. (2010) "Using student test results for accountability and improvement." OECD Education Working Paper 54, DOI:10.1787/5km4htwzbv30-en

[e] "Educational Attainment in Rural Areas" A report prepared for the Department for Environment, Food and Rural Affairs by the National Centre for Social Research (NatCen), 31 December 2009. http://www.natcen.ac.uk/media/665690/c5974000-7879-4f01-890f-c31cf9ca7489.pdf

[f] "Fair Admissions to Higher Education: Research to describe the use of contextual data in admissions at a sample of universities and colleges in the UK." Research report by Kath Bridger, Jenny Shaw (BSV Associates Ltd) and Joanne Moore (ARC Network) for the Supporting Professionalism in Admissions (SPA) Programme. http://www.spa.ac.uk/documents/ContextualData/Full_SPA_Contextual_data_Research_Report-Feb2012.pdf

[g] Head of Quantitative Analysis for Policy, HEFCE.
May be contacted to corroborate that the most recent versions of MLwiN were needed to carry out the analyses.

[h] "Selection of Staff for inclusion in RAE2008,"
http://www.hefce.ac.uk/media/hefce1/pubs/hefce/2009/0934/09_34.pdf