

NCRM Nodes Phase 2: 2008-2011

ADMIN: Administrative data: methods, inference and network **L. Dearden, Institute of Education**

Policy-oriented research is increasingly making use of administrative data sets in a range of fields including welfare-to-work, employment, health, and education. The great strength of administrative data is that they have information on almost everyone. However, the main weakness is that they often lack important variables; for instance, the National Pupil Database (NPD) has detailed information on children's educational outcomes but no information on the parent's education or employment status. In contrast, survey data provide very rich information. Clearly, there is scope for using survey data to enhance administrative data. An important weakness of survey data is that, unlike administrative data, they include only a sample of individuals and suffer from problems such as non-response, attrition, and recall bias. Clearly, there is scope for using administrative data to enhance survey data.

The main aim of ADMIN is to develop and disseminate methodology for making best use of administrative data by exploiting survey data, and for making best use of survey data by exploiting administrative data. The Node will apply this methodology to deliver high-quality evidence on some key policy questions. The Node therefore seeks to add significantly to both the methodological and policy research agenda. Another central aim of the Node is to build capacity for methodologically strong research that makes good use of administrative and survey data to inform public policy in the UK. Specifically, the objectives of this node are to:

- Develop innovative methodological approaches for using survey data to enhance methods of analysis of administrative data. This includes developing methods for making inferences when covariates or responses are missing for non-survey respondents in the linked data, and developing variance estimation tools when such variables have been filled in by imputation when multiple imputation is not feasible.
- Develop innovative methodological approaches for using administrative data to enhance methods of analysis of survey data. This includes using administrative data to evaluate and overcome problems of measurement error in survey variables such as recalled event histories and ethnicity, and bias due to attrition in longitudinal surveys. Methods for using administrative data to improve small-area estimates of the means and quantiles of survey variables will also be developed.
- Build capacity in the quantitative research field via training courses, web based materials, fellowships and studentships. In particular the Node will develop a range of high quality courses on data analysis methods, with particular focus on methods for analysing linked administrative / survey data.
- Develop an international network of researchers focused on the exploitation of linked administrative / survey data.
- Contribute to substantive policy research questions in the field of education and learning, such as the extent of and constraints on school choice and how one best measures school effectiveness and pupil progress, taking account of factors outside the school's control.

Bayesian methods for integrated bias modelling and analysis of multiple data sources in observational studies (BIAS II)

N. Best

The activities we propose to focus on in BIAS II are outlined below:

Methodological development – this will be motivated by the methodological work of BIAS I, with the aim of (i) expanding the promising developments in modelling multiple biases in observational data to encompass additional types of biases and new approaches for dealing with them, with reference to non-response in surveys, (ii) to extend Bayesian modelling of small area data to include data measured over time. We further include (iii) a new focus on developing models for combining multiple longitudinal data sets. Various substantive applications in health and social science will motivate this work, including some arising from our continuing collaboration with ONS, and some from new collaborations with psychologists in the USA, geographers in Cambridge and Canadian statisticians interested in multiple bias modelling.

Applications of BIAS I methodology – The hierarchical related regression (HRR) methodology for combining individual and aggregate data that we successfully developed in BIAS I has the potential to be useful in a variety of areas outside the epidemiological contexts we have so far considered. We have identified two groups of collaborators, one in political science (Oxford) and one in occupational health (Finland), with whom we will work to apply HRR to substantive problems.

Training and Capacity Building – to help facilitate a step change in the skill base of quantitatively minded social scientists, we will expand our existing workshops on Bayesian modelling to include more problem-based learning and participant-led activities, provide a new mentoring scheme offering one-to-one support in use of Bayesian methods, develop web based training material (collaboration with Bristol node) and free software to implement methods we develop. The uptake and dissemination of the training material will be facilitated by the active participation in our training activities of the wide range of social science collaborators that we have established.

The Lancaster-Warwick-Stirling Node: Developing statistical modelling in the social sciences phase 2

B. Francis, University of Lancaster

The aim of the node in our original application will also be our aim for phase 2 of the NCRM nodes; namely to develop and extend statistical methodology and models related to correlated and longitudinal data with substantive applications in criminology, psychology and sociology, and to provide an excellent programme of training and capacity building to support researchers in this area.

The five original objectives will remain and a new objective 6 is explicitly added to underline our commitment to collaboration within the Centre:

1. To develop and extend statistical models for correlated and structured longitudinal data, and associated computational methodologies, building upon existing expertise and experience.
2. Using such models, to investigate relevant and topical social science problems in criminology, psychology and sociology and to seek cross-fertilisation in these areas and into other disciplines.
3. To develop software and training materials to allow such models to be disseminated to other users.
4. To provide a programme of short courses and workshops in the broad area of statistical modelling and advanced courses related to node activity.
5. To liaise with regional and national user groups, and to develop international links with other research groups in this area.
6. To liaise with other nodes and the hub in organising joint research and capacity building activity.

STRUCTURES for building, learning applying and computing statistical models (Lemma 2)

J Rasbash, University of Bristol

Research

The following social science research questions will be addressed:

- Is the UK education system becoming more segregated over time and space?
- How large are family effects compared to school and pupil effects on pupils' learning progress?
- Is there any evidence for school competition affecting pupils' learning progress?
- Does parental selection into neighbourhoods and schools drive the school competition process?
- How do family risks such as marital conflict, parental depression and children's emotional and behavioural problems interact?

Data from the Avon Longitudinal Study of Parents and Children (ALSPAC), the Avon Brothers and Sisters Study (ABSS) and the Pupil Level Annual Schools Census (PLASC) will be used to address these questions. The first two data sets were collected at Bristol and Bristol runs the ESRC-funded PLASC/National Pupil Database user group (PLUG). We therefore have extensive in-house expertise on these data sets. To better address these and other substantive questions, we will develop new methodology for handling missing data, non-independence within and between random classifications and correlated processes.

Training and capacity building

We will extend the training and capacity building system developed under LEMMA. Colleagues in the Graduate School of Education, Baird and Barnes, have expertise in general models of learning and we will hone these for the evaluation and further development of our training materials. An important aspect of our epistemological framework is that frameworks to support learning techniques can not be separated from learners' socio-cultural contexts. For the Node to be truly successful, it will need to motivate, develop and support a wider user community. We have developed a model of the user community of practitioners of quantitative social science which provides a framework for our proposed developments in applied social science research, methodology, TCB and software. An integrated software infrastructure is essential to support the proposed developments in new methodology, and to provide a suitable modelling environment for cutting-edge applied research as well as TCB. We will develop a new type of statistical software environment targeted at a diversity of users that will provide the platform for a new generation of quantitative social scientists from the novice consumer to the highly experienced developer. This is highly ambitious, but builds on the firm foundation of previous work by ourselves and others, including extensive prototyping.

Realities (Real life methods for researching relationalities)
J Mason, University of Manchester

The proposed Realities Node grows from, develops and extends the work of the current Real Life Methods Node (RLM) of the NCRM. Its overall aim is the pioneering development, use and innovation of 'real life methods' for researching relationalities and personal lives in complex worlds. Realities will move the distinctive methodological field of real life methods into a new and exciting phase, as well as securing and consolidating the considerable resources and achievements of RLM. Realities' aims for achieving a step change in UK social science methodologies are:

- To pioneer and develop approaches and methods to tackle the challenges of researching, theorising and comprehending the significance of relationalities.
- To develop and extend the capacities of the UK social science community to produce compelling data about and insight into the multi-dimensional realities of personal life.
- To develop novel and interdisciplinary research questions that push the boundaries of established ways of knowing, through a question-driven investigative epistemology for researching relationalities in real lives.
- To develop and disseminate ways of combining 'qualitative' and 'quantitative' methods and analysis, that exploit the distinctive strengths of each set of approaches. Through these activities Realities will take up and extend the role that RLM has started to play in helping to transcend the unhelpful elements of the so-called qualitative-quantitative divide.
- To develop and disseminate methods through which we can explore, grasp and explain the significance of 'tangibles and intangibles' in real lives.
- To develop analytical strategies, and ways of thinking and theorising with data, that confront the challenges of producing explanations that are simultaneously evocative (with a real life resonance), rigorous, and generalisable. Here we will further develop work begun in RLM on strategies of 'cross contextual' and 'dialogic' explanation, that critically engage different logics of generalisation in a dialogical manner

SIMIAN: Simulation innovation
N. Gilbert, University of Surrey

SIMIAN has four linked objectives under the general theme of innovation and social change:

1) To advance the methodology of simulation and thus its substantive applicability by developing three "demonstrator" simulations in areas chosen to be of relevance across the social sciences. These are modelling behaviour in repeated strategic interaction, modelling social responses to genuine novelty (innovation) and modelling the operation of norms in social systems. As well as their substantive relevance, each of these demonstrator systems has been chosen to address a recognised challenge to simulation methodology: synthesis of competing theories to facilitate progressive research, simulating systems without pre-defined categories and synthesising data from different measurement techniques respectively. The outcome of this objective will be innovative simulations that contribute both to the methodology of the technique in important ways and to substantive debates in the social sciences.

2) To identify and respond to the needs of both existing and potential adopters of simulation across the social sciences. The outcome of this objective (carried out as part of other activities) will be an effective plan for dissemination and TCB based on the identification of different constituencies, combined with a "roadmap" for the development of future research, training and capacity building.

3) To provide TCB activities appropriate for the different constituencies identified above.

These will include (but are probably not limited to):

- Direct production of good quality research to convince the open minded but sceptical within the academic community that simulation has something concrete to contribute. For a new method, it is as important to have a positive attitude from non-adopters as to enthuse and train adopters.
- A systematic programme of training activities to take those who are already interested from novice to expert (as appropriate).
- "Meeting of Minds" and distinctive TCB activities (mentoring, consultancy, user fellowships) to facilitate dissemination to the non-academic community, recognising diversity of goals and expertise in different communities.
- A programme of advanced training and resource development to reduce the costs of disseminating simulation in social science. For example, training to support academics who want to supervise doctorates in simulation or write effective
- funding proposals and online resources to allow self-study and make it easier to run courses in simulation as part of existing undergraduate and post graduate training programmes. The outcome of this objective will be a TCB framework that allows simulation to reproduce itself in the social sciences at a suitable level.

4) To create capacity in UK simulation that will continue after the initial funding period but which could not have been produced without funding. The outcome of this objective will be collaborative research networks, specific training activities (like workshops) which, once created, can be delivered at relatively low cost subsequently, increased human capital in the social science community, a more accurate and positive view of simulation in the social sciences (particularly among non-academics and non-adopters) and so on.

Qualitative innovations in CAQDAS (QUIC)

N. Fielding, University of Surrey

The proposed Node is informed by increasing convergence in digital technologies available to support social research. Recent affordances enable significant enhancement of the functional integration of methodologies and computational resources in (i) the integration of quantitative and qualitative data; (ii) the systematic analysis of multi-stream visual data; and (iii) the convergence of geo-referencing technologies and methodologies with qualitative software. The Node aims to capture, contribute to, and bring to the mainstream, cutting-edge integrative developments in these fields.

The Node's principal strands are (1) Methodological Innovations in Computational Support ('MICS') for qualitative and mixed method research, and (2) a Training and Capacity Building ('TCB') programme to raise awareness of, and skills in, fundamental and advanced qualitative software use. The main MICS aim is to apply, evaluate and document the emergent developments in digital technologies and methodologies noted above. The main TCB aim is to bring these technologies and methodologies into the mainstream of research practice.

MICS will develop, apply, evaluate and document emergent methodologies via demonstrator projects in the substantive field of social aspects of environmental risk - natural risk arising from climate change, and human-made risk arising from crime and social disorder. TCB will draw on methodological and technical experience resulting from the MICS strand, and from long experience in training qualitative software ('CAQDAS') users, to support the 'digitally convergent' researcher in three ways: by developing interactive online protocols to support users in planning and undertaking projects and data analysis using the emergent methodologies and computational resources subject of the MICS strand; by developing web-based research archives and teaching materials derived from the MICS strand; and by convening seminars as awareness-raising and methodological reflection events on a UK outreach basis capitalising on the structure of the NCRM.

The main aims of the Node's two strands embrace several component objectives. The MICS data integration stream will apply emergent computational support for methodologies involving configuration of qualitative data for statistical analysis. The computational resources and techniques will be evaluated comparatively in a secondary analysis of factors affecting public response to climate emergency warnings. The MICS visual data analysis stream builds on pioneering work in using the Access Grid for 'virtual fieldwork'. It will conduct AG-based fieldwork with Environment Agency policy and field officers, with a prime objective of evaluating AG utility for applied research in real-world settings, and subsidiary objectives including demonstrating a carbon neutral means of conducting research to applied audiences in environmental risk management. The MICS geo-referencing stream will innovate by incorporating GPS technology into a tool used by police and local authorities in 'environmental scans' of crime and disorder risk perceptions and by evaluating the provision of Google Earth functionality in a leading CAQDAS package, with which the output from the environmental scan tool will be analysed.

The MICS streams share the objective of documenting their focal methodologies in such a way as to enable the TCB to transfer experientially-based knowledge about their use to current and next-generation researchers. By raising awareness of new approaches to the collection, analysis, management and curation of qualitative and mixed method research data the Node will make more widespread the use of techniques currently practiced only by the most advanced users. While the prime target for the Aims and Objectives is the social science research community, the substantive focus on environmental risk opens avenues to significant policy and research-user audiences.