Evidence and Evaluation in Social Policy
The object of this series, in this age of re-thinking on social welfare, is to bring fresh points of view and to attract fresh audiences to the mainstream of social policy debate.

The choice of themes is designed to feature issues of major interest and concern, such as are already stretching the boundaries of social policy.

This is the seventeenth collection of papers in the series. Previous volumes include:

- Crime and Social Policy *H. Kemshall*
- The Times They Are Changing? Crisis and the Welfare State *B. Greve*
- Reforming Long-term Care in Europe *J. Costa-Font*
- Choice: Challenges and Perspectives for the European Welfare States *B. Greve*
- Living in Dangerous Times: Fear, Insecurity, Risk and Social Policy *D. Denney*
- Reforming the Bismarckian Welfare Systems *B. Palier and C. Martin*
- Challenging Welfare Issues in the Global Countryside *G. Giarchi*
- Migration, Immigration and Social Policy *C. Jones Finer*
- Overstretched: European Families Up Against The Demands of Work and Care *T. Kröger and J. Sipilä*
- Making a European Welfare State?: Convergences and Conflicts over European Social Policy *P. Taylor-Gooby*
- The Welfare of Food: Rights and Responsibilities in a Changing World *E. Dowler and C. Jones Finer*
- Environmental Issues and Social Welfare *M. Cahill and T. Fitzpatrick*
- The Business of Research: Issues of Policy and Practice *C. Jones Finer and G. Lewando-Hundt*
- New Risks, New Welfare: Signposts for Social Policy *N. Manning and I. Shaw*
- Transnational Social Policy *C. Jones Finer*
- Crime & Social Exclusion *C. Jones Finer and M. Nellis*
CONTENTS

List of Contributors vii

Introduction: Evidence and Evaluation in Social Policy 1
Ian Greener and Bent Greve

1 Trials and Tribulations: The ‘Use’ (and ‘Misuse’) of Evidence in Public Policy 5
Christopher Deeming

2 Understanding the Influence of Evidence in Public Health Policy: What Can We Learn from the ‘Tobacco Wars’? 29
K. E. Smith

3 Caught in the Same Frame? The Language of Evidence-based Policy in Debates about the Australian Government ‘Intervention’ into Northern Territory Aboriginal Communities 47
Emma Partridge

4 A Systematic Review of Comparative Studies of Attitudes to Social Policy 63
Trude Sundberg and Peter Taylor-Gooby

5 Public Opinion and Policy-making 81
Ray Pawson and Geoff Wong

6 Obstacles to Evidence-based Policy-making in the EU Enlargement Countries: The Case of Skills Policies 97
Will Bartlett

7 Understanding Employment Barriers for Lone Parents in Great Britain: Research Gaps and Missed Opportunities 115
Tina Haux

8 Putting the Research Boot on the Policymakers’ Foot: Can Participatory Approaches Change the Relationship between Policymakers and Evaluation? 129
Liz Richardson

Index 147
List of Contributors

Will Bartlett, European Institute, London School of Economics and Political Science, London, UK.

Christopher Deeming, Geographical Sciences, University of Bristol, Bristol, UK.

Ian Greener, Durham University, Durham, UK.

Bent Greve, Roskilde University, Roskilde, Denmark.

Tina Haux, University of Lincoln, Lincoln, UK.

Emma Partridge, Social Policy Research Centre, University of New South Wales, Sydney, Australia.

Ray Pawson, School of Sociology and Social Policy, University of Leeds, Leeds, UK.

Liz Richardson, School of Social Sciences, University of Manchester, Manchester, UK.

K. E. Smith, Global Public Health Unit, School of Social and Political Science, University of Edinburgh, Edinburgh, UK.

Trude Sundberg, University of Kent, Kent, UK.

Peter Taylor-Gooby, University of Kent, Kent, UK.

Geoff Wong, Centre for Primary Care and Public Health, Queen Mary, University of London, London, UK.
What counts as evidence in social policy, and how evidence does, or perhaps more often, does not, influence policy-making, have become central questions in the last 20 years. The evidence-based medicine movement gathered momentum in the 1980s and 1990s (Sackett et al. 1997), creating a framework for the assessment of research in that field and how it might lead to a more robust basis for clinical decision-making, and even health policy. This led to policymakers, especially the New Labour government in the UK, suggesting that the ideological and interest-based politics of the past were now to be jettisoned in favour of an approach that was based instead on a pragmatic, ‘what works’ basis instead (Davies et al. 1999).

Still, in times of austerity, policy suggestions that are based on solid evidence, especially where they offer opportunities for efficiency savings, will, all other things being equal, have a higher chance of being accepted. The counter-argument naturally being that this also implies a risk that policymakers will only use evidence pointing in the same direction as their existing biases, especially given that evidence is seldom straightforward in its interpretation to policy ends, and this is likely to result in policymakers wanting to be cautious and so tending towards policies that are based on conclusions they feel most comfortable with.

This last issue is especially prevalent in many areas of social policy services where the relation between user and producer, at least to a certain extent, is still a kind of black-box where it can be difficult from the ‘outside’ to see exactly what is going on, and so to track how policy is being influenced. However, this does not reduce the need for knowing more about what does and does not work in social policy as a way of not only using scarce resources in the best way possible, but also to be able to give users the best possible type of care.

From the perspective of 2013, looking back at a decade of evidence-based policy-making, we should be able to see clear examples of improved policy-making through an increased use of evidence, and a transformation in which research and evaluation drive the policy process, alongside methodological innovation and closer working between researchers and policymakers. Is this actually the case?

The chapters in this book, which were responses to a call asking researchers to specifically address the role of evidence and evaluation in policy-making, suggest a less rosy picture than we might hope.
Chris Deeming’s chapter considers the role of randomized controlled trials (RCTs) in shaping public policy, exploring why RCTs have become regarded as the gold standard in many low- and middle-income countries, but not necessarily in the richer, liberal democracies. Deeming also considers a range of issues associated with the use of RCTs in social policy, including ethical and methodological issues, and the use and misuse of evidence and evaluation in social policy. It is a good introduction to the full range of issues covered in the book, as well as making a contribution in its own right through its comparison of the use of RCTs in the global north and developing south.

Katherine Smith’s chapter on the ‘tobacco wars’ suggests that, rather than being based on research, policy-making with respect to public health and tobacco is perhaps better explained using the methods of political science, in which coalitions and interest groups work to secure their demands, and that evidence is enrolled to those ends rather than driving the process. Nevertheless, one could argue that without many years of substantive evidence on the link between smoking and health, changes would not have been possible.

From an entirely different context, Emma Partridge presents the case of policy-making from the Northern Territory of Australia where the language of evidence-based policy has been used by both the government and its critics. Similarly to Smith’s work, in some respects, Partridge emphasizes the political basis of the relationship between knowledge, evidence and policy-making. More than this, in Partridge’s view, using evidence-based policy as a prescriptive ideal can make policymakers lose sight of other possible strategies for improving policy towards indigenous people.

Trude Sunberg and Peter Taylor-Gooby’s chapter calls into question the idea that systematic reviews can straightforwardly inform policy-making, suggesting that they require considerable care in their use because of potential biases toward US-based research, and the weaker reporting of book-based research compared to articles. Where secondary research is based on research summary and review, this can be a quick and effective way of generating evidence – but, as the authors demonstrate, it is not without significant problems.

Ray Pawson has been at the centre of methodological innovation with respect to the review and evaluation of social policy since the 1990s. Here, his chapter with Geoff Wong considers many of the concerns of Smith, Sunberg and Taylor-Gooby’s work in considering the problems of measuring public support for policy interventions. Pawson and Wong suggest that systematic reviews can be useful, but that we need to take a more considered approach to measure not only snapshots of public opinion, but also how opinion is formed. To illustrate their approach, the authors explore public support for legislation banning smoking in cars carrying children. They show how public opinion data can supplement or even supplant a limited scientific research base through a diachronic assessment of public opinion formation that suggests that public opinion can play a wider role in policy-making than is often assumed to be the case.

Turning to a different set of challenges, Will Bartlett’s chapter considers skills policies in EU enlargement areas, but even more saliently, the consequences of the economic crisis on countries in periods of rapid change for
policy-making more generally. He suggests that, in these circumstances, state capture and a coercive form of policy transfer occur that are a considerable distance from models of evidence-based policy-making. This apparently depressing story, however, can represent an opportunity for research to play a greater role in policy-making, especially for the evaluation of government programmes, but in a context where advocacy coalitions are a central feature of the political process.

Tina Haux’s chapter is an illuminating case study of the relationship between evidence, research and social policy, considering work commissioned by the UK Labour government between 1997 and 2010 to explore the employment barriers for lone parents not in work. Haux finds that research often appears to have been commissioned in areas where it was not taken up by policy, but also that significant gaps occurred where policymakers needed research which was not commissioned. Even though, under the post-2010 Coalition government, policy has ‘intensified’, continuing and expanding many of Labour’s ideas, there is still a lack of robust data and research supporting the government’s approach, while at the same time the main research vehicle for monitoring changes to the composition and outlook of lone parents has come to an end, giving us incomplete data series. All in all, this is not an encouraging picture.

Lastly, Liz Richardson’s chapter presents an innovative perspective in exploring how local government decision-makers can engage more fully with evaluation research – through the use of participative and co-productive methods. The research finds that these methods helped increase local politicians’ awareness of their own ideological biases, and even an enhanced capacity for learning. It is important to note, however, the policymakers in the case study struggled to achieve robust research designs and methods, which may in turn have actually increased their appreciation of the value of academic expertise (something which we can enthusiastically endorse). Richardson suggests that politicians’ belief in positivism does not have to be a barrier to more participative evaluation, and offers helpful suggestions for how this can be more successfully achieved. Her work offers the special edition an upbeat ending, showing the potential for bridging the gap between research and policy-making.

The work covered in the special edition, then, presents a very mixed picture of the role of evidence and evaluation in social policy in 2013. Politics remains central to decision-making, and in an environment of continued fiscal austerity, the opportunities for conducting funded research for central government are diminished. There is potential for the synthesis of existing research, but such syntheses need to be treated with care and caution rather than simply trying to import methods from systematic reviews and RCTs in scientific and clinical work. However, these kinds of reviews are often also useful in attempting to systematize what we know at least something about and what we do not know anything about, as well as highlighting areas where there is a stronger need for more research.

Naturally, we also need to take greater account of the specific policy and economic context when exploring the role of evidence in policy-making, and trying to assume such variables away does little to increase the persuasiveness
or potential to work of our research. Although this is not a counter-argument of using evidence, it is more a word of caution in the way one can use evidence. This also implies a stronger need for students of social policy and administration to know how and to what degree certain types of research are conducted, and how one can substantiate an argument in order to avoid stating that ‘anything goes’. Solid knowledge and information are thus clearly still extremely important in order to ensure the best possible service for the users of the welfare state.

Lastly, Richardson’s chapter especially stresses that it is possible for researchers and politicians to work together, but that it will involve compromises on both sides, and that smaller-scale, more localized work might be the way forward. At a time when funded opportunities in Europe appear to be moving towards large-scale, mega-projects, this is a salutary lesson indeed.

References

Trials and Tribulations: The ‘Use’ (and ‘Misuse’) of Evidence in Public Policy

Christopher Deeming

Introduction

We can achieve a sort of control under which the controlled, though they are following a code much more scrupulously than was ever the case under the old system, nevertheless feel free. They are doing what they want to do, not what they are forced to do. That’s the source of the tremendous power of positive reinforcement – there’s no restraint and no revolt. By careful cultural design, we control not the final behaviour, but the inclination to behave – the motives, desires, the wishes. (Skinner 1948: 246–7)

Many observers have commented on the spread of behavioural experiments in global policy. This new and emerging form of public policy often involves conditioning the receipt of welfare – in the form of cash transfers, goods or services – according to specific individual behaviour. Few however, have touched on the contextual differences in terms of the application of research and evidence governing policy, especially social assistance in the developing world and social security within the developed world. Subsequently, the debates on the progress of public policy in the global north and south are not as well connected as they might be.

In LMICs, social experiments with conditionality are part of the drive for evidence-based policy (Fiszbein et al. 2009). Conditional cash transfer (CCT) programmes usually have an *a priori* evaluation design built into their operation, that embraces experimental or quasi-experimental features and RCT designs, for example. By contrast, in the advanced liberal democracies, there has been less direct appeal to research evidence gained using from robust evaluation in order to secure major welfare reform: experimentation and evaluation with RCTs has been less of a priority. Consequently, welfare states were reformed on ideological grounds, with an appeal to political theory as Mead and Beem (2005) observe. Putting this in stark terms, welfare conditionality in the south is, arguably, being driven by an evidence-based policy-making agenda, whereas in the north, political philosophy is clearly driving...
welfare reform. This chapter seeks to shed new light on the relationship between evidence and evaluation within the different worldly contexts, by drawing out emerging arguments and counter-arguments about the ‘use’ and ‘misuse’ of evidence within public policy. The chapter is organized as follows: the first section examines the increase and nature of evidence-based policy for development; the second section considers how welfare policy has been transformed in the developed world; followed by a more detailed examination of some of the controversies in section three. Lastly, reflections and conclusions are drawn together in the fourth section.

Public Policy for Development: The Rise of Behavioural Economics

In the brave new world of the behavioural economist, achieving the Millennium Development Goals (MDGs) for health and well-being is ultimately about demonstrably changing people’s behaviour for the better. Being able to demonstrate the effectiveness or impact of an intervention is thus the keystone for policy development. As a result of the global research effort, it is now well established that there are certain (desirable) human behaviour and conditions of living that are beneficial for our health and well-being that, arguably, should form the basis for public policy (Dean 2010). This point, implicit in the literature, certainly needs to be made more explicit, as biomedical and social research continues to establish food and dietary requirements for good health, education, housing and living standards, along with certain behaviour and practices that reduce the risk of ill-health and disease (e.g. WHO 2002). For far too long, the criticism has been that acceptance of incontrovertible evidence into functioning policy has been slow, partial and unsystematic; resulting in health deficits, waste of human potential and other associated costs to society (WHO 2011). The second point, about clearly demonstrating success and impact in social policy addresses the need for robust evaluation. In order to consider the effectiveness of a particular policy or programme, one really needs to know what outcomes would have been achieved had the programme not been in place. This is often referred to as the ‘counterfactual outcome’ (OECD n.d.). One way of overcoming this evaluation problem is through the use of RCTs, which are considered to be the ‘gold standard’ of all the methods available to researchers (Young et al. 2002). In the trials, participants are usually randomly assigned to intervention or control groups (Sibbald and Roland 1998), researchers then compare the outcomes between groups (see figure 1.1). A systematic review of RCT results can usually be found at the very pinnacle of the research evidence hierarchy (figure 1.2), as it provides a way of pooling evidence from different studies to provide an overview of outcomes (White and Waddington 2012).1

It may not be surprising, therefore, to find an influential group of economists advocating social experiments and in particular RCTs, as the main tool for studying the effectiveness of policy in development settings. In a recent trial in Malawi, for example, poor families were given cash on the condition that they send their children to school (Baird et al. 2000a, 2000b). Some villages were randomly assigned to the social assistance programme and
Importantly, at the outset, each village had the same equal chance of being picked to receive the intervention; families were not told that they were part of an experiment, they were ‘blind’ or ‘blinded’: knowing this may affect their behaviour and the outcome of the study. The development economists – sometimes referred to as the ‘randomistas’ – argue that this type of randomized experiment is the only sure way of identifying impact, because it helps to eliminate bias and other confounding (hidden) factors. Other non-experimental methods found at the bottom of the evidence hierarchy are largely dismissed as they are considered unscientific and are best avoided (figure 1.2). The influence of the randomistas appears to be growing; non-governmental organizations (NGOs) (the World Bank, the International Monetary Fund, the World Trade Organization), philanthropic agencies and donors (e.g. DFID 2011b) are increasingly giving (explicit) preference to randomized designs and systematic reviews in evaluating public policy programmes and their impacts (Hickey et al. 2009; Hagen-Zanker et al. 2012).

According to behavioural economists like Banerjee and Duflo (2011), the major debates in international development can be boiled down to disagreements about the shape of a function in development theory (figure 1.4). The S-shaped curve on the left suggests poor people are ‘trapped’ in poverty and require a ‘conditional push’ to get out of their own poverty. The L-shaped curve on the other hand, suggests that poor people are gradually able to pull themselves out of poverty because they are not really ‘trapped’ at all. If we

---

**Randomized controlled trials (RCTs)**

Randomized controlled trials (RCTs) are the most rigorous way of determining whether a cause-effect relation exists between treatment and outcome and for assessing the cost effectiveness of a treatment. They have several important features:

- Random allocation to intervention and control groups.
- Participants and trialists should remain unaware of which treatment was given until the study is completed (although such double blind studies are not always feasible or appropriate).
- All groups are treated identically except for the experimental treatment or intervention.
- The analysis is focused on estimating the size of the difference in predefined outcomes between intervention and control groups.

Other study designs, including non-randomized controlled trials, can detect associations between an intervention and an outcome. But they cannot rule out the possibility that the association was caused by a third factor linked to both intervention and outcome. Random allocation ensures no systematic differences between intervention groups in factors, known and unknown, that may affect outcome.

*Source:* adapted from Sibbald and Roland 1998: 201.

others were not (figure 1.3). Importantly, at the outset, each village had the same equal chance of being picked to receive the intervention; families were not told that they were part of an experiment, they were ‘blind’ or ‘blinded’: knowing this may affect their behaviour and the outcome of the study. The development economists – sometimes referred to as the ‘randomistas’ – argue that this type of randomized experiment is the only sure way of identifying impact, because it helps to eliminate bias and other confounding (hidden) factors. Other non-experimental methods found at the bottom of the evidence hierarchy are largely dismissed as they are considered unscientific and are best avoided (figure 1.2). The influence of the randomistas appears to be growing; non-governmental organizations (NGOs) (the World Bank, the International Monetary Fund, the World Trade Organization), philanthropic agencies and donors (e.g. DFID 2011b) are increasingly giving (explicit) preference to randomized designs and systematic reviews in evaluating public policy programmes and their impacts (Hickey et al. 2009; Hagen-Zanker et al. 2012).

According to behavioural economists like Banerjee and Duflo (2011), the major debates in international development can be boiled down to disagreements about the shape of a function in development theory (figure 1.4). The S-shaped curve on the left suggests poor people are ‘trapped’ in poverty and require a ‘conditional push’ to get out of their own poverty. The L-shaped curve on the other hand, suggests that poor people are gradually able to pull themselves out of poverty because they are not really ‘trapped’ at all. If we