# Policy Evaluation in the United Kingdom

Philip Davies PhD

Government Chief Social Researcher's Office
Prime Minister's Strategy Unit
UK Cabinet Office
London
SW1A 2WH
England

Paper to be presented at the
KDI International Policy Evaluation Forum,
Seoul, Korea
May 19-21, 2004

**Abstract**

The United Kingdom Government uses a wide range of evaluation methods to ensure that policies, programmes and public services are planned and delivered as effectively and efficiently as possible. A major driving force for high quality policy evaluation in the U.K. is the Government's commitment to evidence-based policy making. This requires policy makers, and those who implement policies, to utilise the best available evidence from national statistics, academic research, economic theory, pilots, evaluations of past policies, commissioned research and systematic consultation with delivery agents. The Government's strategy for public spending and taxation also provides the context within which policy evaluation takes place in the U.K.

The paper reviews the types of evaluation that are used by the UK Government, including impact evaluation, implementation evaluation, economic evaluation, and the use of descriptive and inductive statistics for evaluation purposes. The use of Performance Management for the allocation and accountability of resources by the UK Government is described, as is the machinery that has been developed in the UK to deliver better public services.

The paper concludes by considering the role of factors other than evidence and policy evaluation in the UK policy making process.

.

## Introduction

The United Kingdom Government uses a wide range of evaluation methods to ensure that policies, programmes and public services are planned and delivered as effectively and efficiently as possible. This includes the use of systematic reviews of existing evidence (Cabinet Office 2003a), policy pilots (Cabinet Office, 2003b), demonstration projects (Morris *et al*, 2004), various *ex ante* and *post hoc* evaluations of specific interventions, economic appraisal and evaluation methods (HM Treasury, 2003a), strategic audit and international benchmarking (Cabinet Office, 2003c, 2004), regulatory impact assessments (Cabinet Office, 2003d), and performance management mechanisms. This work assists strategic planning and development as well as the operational management and delivery of public services.

A major driving force for high quality policy evaluation in the U.K. is the Government's commitment to evidence-based policy making (Cabinet Office, 1999a, 199b, 2000, 2001b). This requires policy makers, and those who implement policies, to utilise the best available evidence from national statistics, academic research, economic theory, pilots, evaluations of past policies, commissioned research and systematic consultation with delivery agents. The Government's strategy for public spending and taxation also provides the context within which policy evaluation takes place in the U.K. (HM Treasury, 2004)

## Types of Policy Evaluation Used in the U.K.

### Impact Evaluation

*Evaluating Outcome Attainment*

The simplest form of impact evaluation is what is sometimes called a 'goals-based evaluation' (Patton, 2002), in which policy makers want to know whether a desired outcome, target or goal has been achieved. This is a fairly straightforward issue of defining a desired outcome at the outset of a policy initiative and checking at some agreed future time whether this outcome has or has not been achieved. This approach to policy evaluation is used extensively by the UK government and is at the heart of the Performance Management system discussed in greater detail below. One limitation of goals-based evaluations is that they may not also considered the *unanticipated* outcomes or consequences of a policy initiative and, consequently, may give a partial, if not biased, view of the policy's outcomes.

*Evaluating the Net Effect of Policies*

Governments of any country need to know at the outset of policy development, and after policies have been implemented, their likely and achieved impacts in terms of both the positive and negative outcomes. They also need to know the *net effect* of a policy in comparison with doing something else or doing nothing at all (i.e. a counterfactual). Establishing a counterfactual requires a policy model (or theory) that considers realistic options to the policy in question, and a test of these options using experimental or quasi-experimental methods of evaluation.

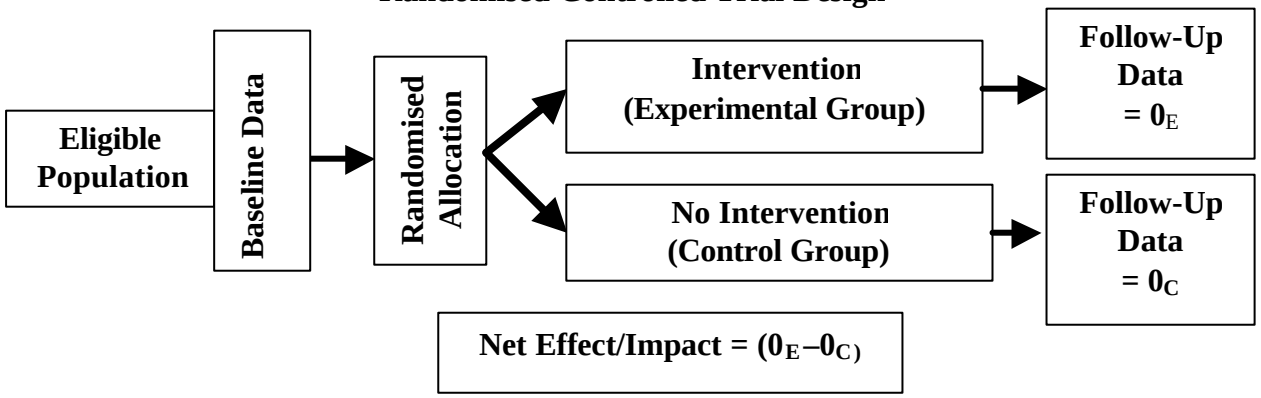
The UK Government is committed to establishing the likely and actual impact of policies using a number of experimental and quasi-experimental designs. These include the following:

*Randomised Controlled Trials*

The most rigorous and robust way of monitoring the net impact of an intervention or programme is the randomised controlled trial (sometimes referred to as a random allocation trial). A randomised controlled trial establishes the net impact of a policy by exposing a group of people (or whole units such as schools, hospitals or geographical areas) to the policy intervention in question (the experimental group) whilst withholding the policy to a comparison group (the control group). The allocation of people or units to the experimental and control group is undertaken on a randomised basis. Baseline data on the anticipated outcomes are collected at the outset on both the experimental and control groups, and again at an appropriate follow-up time. Assuming that the random allocation process is undertaken properly, and that the sample sizes are large enough to identify a *minimum detectable effect*, the difference between the baseline and follow-up measures of outcome represents the net effect of the policy in question. The randomised controlled trial design can be represented as follows:

**Figure 1**

**Randomised Controlled Trial Design**



Although the design of randomised controlled trials is appealingly simple, the implementation and execution of them can be complex and requires considerable operational and analytical expertise. There are usually ethical issues about exposing people (the experimental group) to a policy that may be potentially harmful or, alternatively, of withholding a potentially beneficial policy from another group of people (the control group). In the absence of sound evidence *a priori* that a policy may be beneficial or harmful, however, it is generally considered ethically acceptable to conduct a trial to establish the matter one way or another, as long as one ends the trial as soon valid and reliable evidence has been established. It must also be remembered that in the area of public policy it is not uncommon to introduce initiatives on whole populations without any *a priori* knowledge about their likely positive or negative effects.

The UK Government has undertaken, and is currently undertaking, a number of randomised controlled trials of policy initiatives. In the field of labour market and welfare policy, the Restart evaluation (1990) randomly allocated unemployed people to a compulsory major interview at 6 months unemployment (or no such interview for the control group) to see if this had the effect of successfully reintroducing them to the labour market. This is one of the largest and best-known randomised controlled trials in U.K and it established a clear and positive impact on exits from unemployment with lasting effects still present years later. Other labour market and welfare policies that have been evaluated using random allocation methods include the Benefits Agency Visiting Officer pilot, the New Deal programme for people aged 25 and over, and the New Deal for Lone Parents In-Work Training Grants.

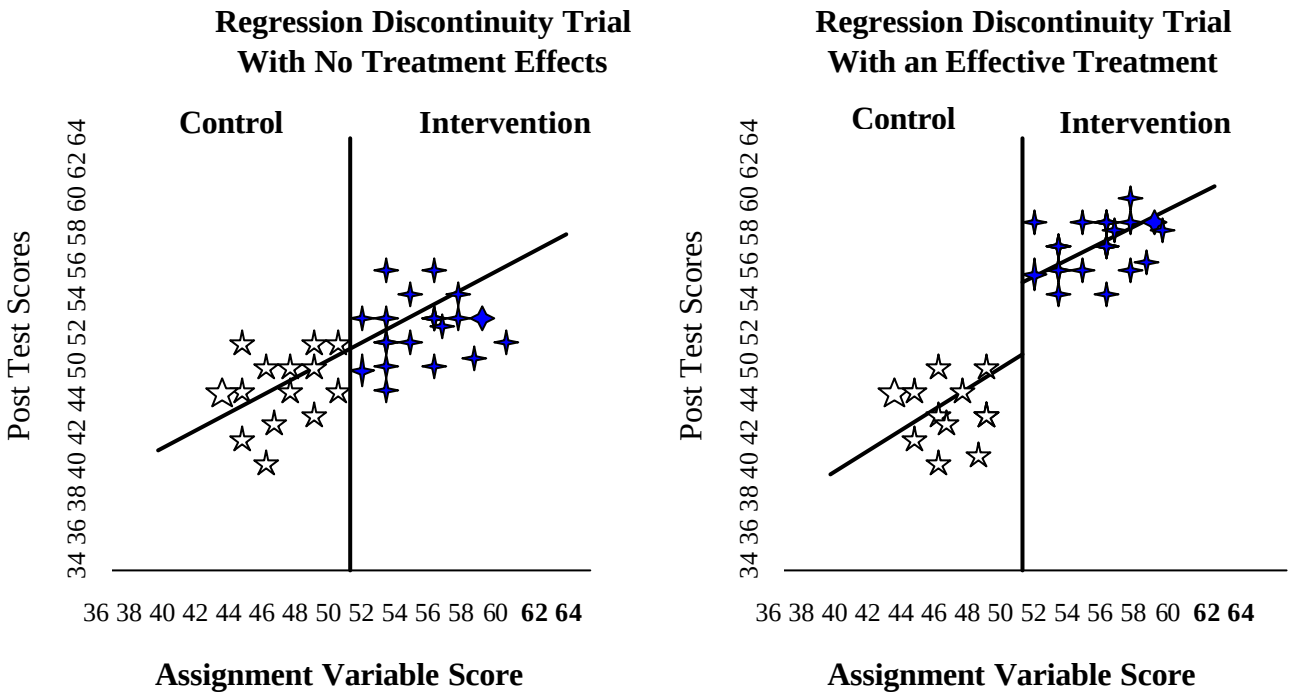
Currently, the UK Department of Work and Pensions in undertaking a randomised controlled trial (the ERA Project) of three policy initiatives aimed at retaining and advancing in the labour market those people on the lower margins of the workforce (Morris *et al*, 2004). The U.K Home Office is currently undertaking a randomised controlled trial (the Restorative Justice Project) in which offenders are randomly allocated to an experimental group, prior to their court appearance, which gives them the opportunity to face their victims and attempt

some degree of restorative justice. The control group does not have this opportunity. The UK Department of Health is currently evaluating the impact of peer-led sex education in UK secondary schools by randomly allocating a group of schools to this programme, and another group of schools to existing ways of teaching sex education without peer-led methods (Oakley, *et al,* 2003). These are just some of the policies or programmes in the U.K. currently being evaluated using a randomised controlled trial design.

*Regression Discontinuity Designs*

Regression discontinuity designs (RDDs) work in a similar way to randomised controlled trials except that instead of allocating people to experimental and control groups on a random allocation basis they are allocated (or assigned) according to a cut off point on a quantifiable pre-test measurement (e.g. some measure of health status or of offenders' risk of re-offending). The assignment variable can be a pre-test on the dependent variable, or can be totally unrelated to the outcome variable. It is important, however, that the assignment variable cannot be caused by the intervention (i.e. it must be independent of the intervention). The regression discontinuity design is most powerful when the cut off is placed at the mean of the assignment variable (Shadish, Cook and Campbell, 2002:209). For each participant in the trial the value of the first measure is plotted on the horizontal axis and the value of the second measure is plotted on the vertical axis. If there is an effect of the intervention or programme it will be detected by a shift (or *discontinuity*) in the regression line representing the pre-test and post-test scores (see Figure 2 below),

**Figure 2**
**Regression Discontinuity Design**



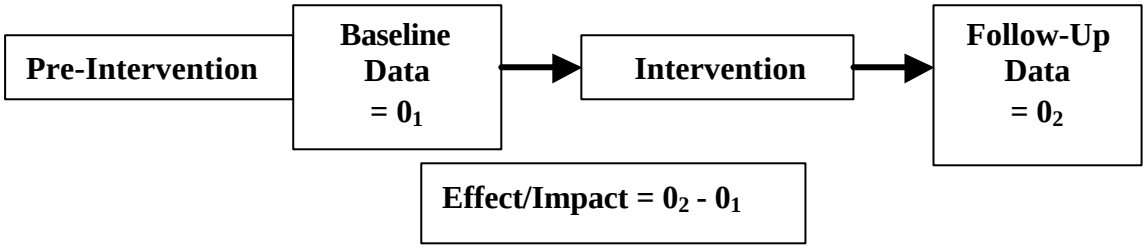

There are very few examples of the use of regression discontinuity designs by UK government evaluators, though a recent UK Home Office evaluation of cognitive behaviour therapy for offenders on probation orders that was initially undertaken using a matched comparisons design (see below) is currently being re-analysed to establish a more precise minimum detectable effect. In this evaluation, offenders were allocated to cognitive behaviour treatment on the basis of their risk of re-offending as measured by a valid and reliable scale (the OGRS scale).

*Single Group Pre- and Post- Test Designs*

The simplest comparative method that is often used in government evaluations is the single group before and after (pre- and post- test) design. This type of evaluation takes a single sample of the population and exposes it to a policy or programme initiative. The sample acts as its own control, and the net effect size is measured in terms of the difference in scores on

the outcome of interest before and after the intervention is introduced. This type of evaluation is represented graphically in Figure 3 below.

**Figure 3**
**Single Group Pre- and Post- Test Design**

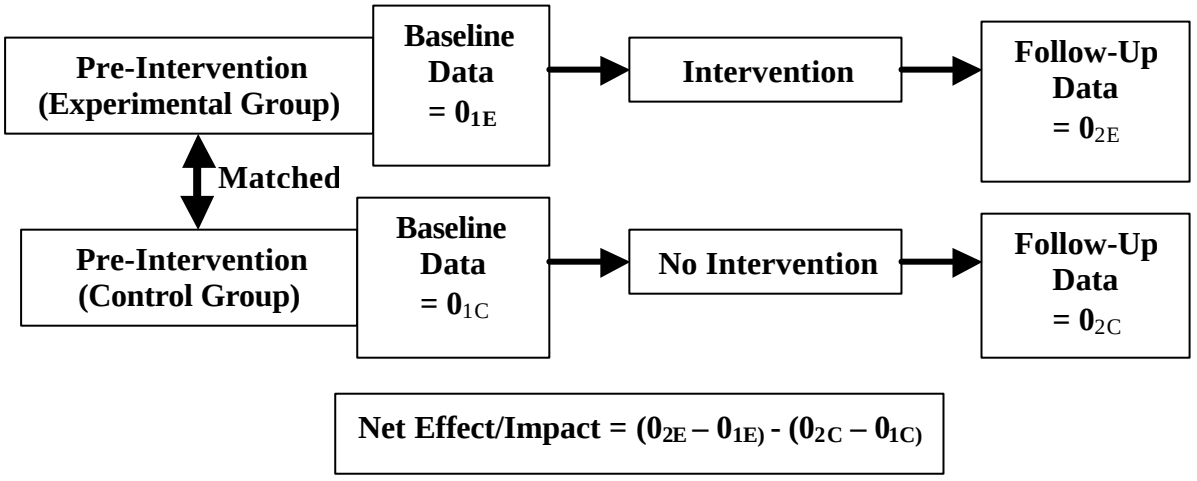



The major weakness of the single group pre- and post- test design is that it cannot control for what would have happened anyway (i.e. in the absence of the intervention) and, therefore, can give a distorted (or biased) measure of the true net impact of the intervention. One way of dealing with this is to use a *matched comparison* design,

*Matched Comparison Designs*

With a matched comparison design an experimental group is exposed to a policy or programme initiative whilst a closely matched control group does not receive the policy or programme in question. Differences between the two groups before and after the intervention on agreed outcomes are used to indicate the net effect of the policy or programme. The comparison group might be a sample of people, a similar unit (such as a hospital, health centre, or school), or an area (such as a housing estate or nearby street).

**Figure 4**
**Two Groups Pre- and Post- Test Design**



The main weakness of the two groups pre- and post- test design is knowing *a priori* on which variables the experimental and control groups should be matched. Typically, variables such as age, sex and ethnicity are used to match samples, but there are often many other potentially relevant variables that might be used. In the absence of valid and reliable *a priori* knowledge of appropriate variables on which to match experimental and control groups the influence of external factors other than the intervention (known as confounds) is unknown, thereby resulting in unaccountable bias.

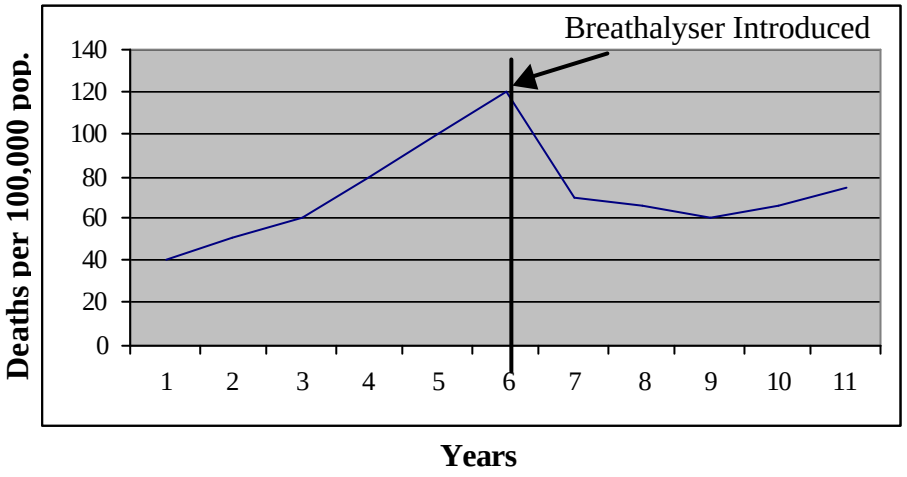
Propensity Score Matching (Rosenbaum and Rubin, 1985) is one method of achieving appropriate matching using a single composite score of the known likely factors that influence the policy outcome of interest. The propensity score on which experimental and control groups are matched is determined after extensive research of the literature on the policy issue in question (e.g. what affects the likelihood of sixteen years old youths staying on or leaving school), and careful multivariate analysis of the relative strength of these contributory factors. The UK Department for Education and Skills used Propensity Score Matching to establish comparable experimental and control groups in order to evaluate the Educational Maintenance Allowance (DfES, 2002). This evaluation was used to test both the most effective way to use payments as an incentive to stay on at school (i.e. whether to make payments to the young people or their parents), and to determine the level of payments that would be most effective. The EMA evaluation gave a very clear message that the most effective way of inducing young people to stay on at school was to make payments to them (i.e. not to their parents) and to provide a level of support between £10 and £30 per week depending on the level of their parents' income. This policy was rolled out nationally in April 2004.

*Interrupted Time Series Designs*

Interrupted time series designs investigate repeated observations of a constant variable over

time and look for 'interruptions' to the series or sequence of observations (see Figure 5). A classic example of an interrupted time series evaluation by the UK Government was the evaluation of the 1967 Road Traffic Act, which introduced alcohol breath tests for drivers of motor vehicles (see Figure 5). Interruptions to time series data might be attributable to a policy intervention, though they could also be a random blip in the series of observations (not likely in the example in Figure 5 where the interruption was continuous). There are different types of change to a time series sequence of observations; changes in the level and in the slop of the curve, the degree of permanency of the effect (as in Figure 5), and the type of impact (immediate or delayed).

**Figure 5**
**Interrupted Time Series Design**
**Road Traffic Fatalities**




In order to attribute the interruption in the time series of observations to a particular intervention it is important to know the specific point (e.g. the date) when the intervention was introduced. This must, of course, have been *before* the interruption in the observations if causality is to be inferred. If this is the case, it is then necessary to consider, and rule out, any other *reasonable* explanations for why the interruption occurred. In the case of the clear

interruption to the time series data on traffic accident fatalities in Figure 5 this would include ruling out, for instance, that there had been a fuel shortage after year 6; or that a major new tax had been introduced on road usage (or petrol); or that there had been a national alcoholic beverages strike; or that the price of alcoholic beverages had increased significantly, and so on. Assuming that none of these alternative explanations is accepted, we can infer that the noticeable reduction in road traffic accident fatalities was almost certainly attributable to the introduction of the breath test for drivers.

It is also important when working with interrupted time series designs to establish that the variable(s) being measured are constant over time. Where definitions and counting practices change frequently over time (e.g. unemployment statistics) it is much more difficult, and sometimes impossible, to use such data as valid measurements, or to establish any causal significance to an interruption in the time series. The UK Government uses interrupted time series designs to evaluate the impact of policies, programmes and projects in areas where process and outcome measures are constant over time, including health care, education, and some crime and justice evaluations.

*Regulatory Impact Evaluation*

Another type of impact evaluation undertaken by the U.K. Government is a Regulatory Impact Assessment (RIA). This is "an analysis of the likely impact of a range of options for implementing a policy change" (Cabinet Office, 2003d). When a policy proposal is being considered, either at the UK domestic or the European level, an RIA setting out all the options under consideration is required.

There are three stages to an RIA; an Initial RIA, a Partial RIA and a Full/Final RIA. A Full/Final RIA should consider the following:

- The *objectives* of the proposed policy.
- The *risks* that the proposal is addressing (which should be quantified). This should consider what effects the policy will have, and on whom, including risks to the environment, consumers, worker safety or health, and business interests.
- The *options* or *alternatives* to legislation such as self-regulation, co-regulation, information and education campaigns, financial incentives, quality marks, recommendation schemes, codes of practice, and doing nothing at all.
- *A costs* and *benefits* analysis of the likely effects on business sectors, including a *Small Firms' Impact Test* and a *Competition Assessment*. The benefits and costs of each option should include those to firms, charities, the voluntary sector, consumers, the public sector, the environment, and the economy at large.
- *Equity* and *Fairness* issues and a *Distributional Impact Analysis*. This should assess the likely impacts of the proposal on particular groups of people, including any transfers of incomes or redistribution of opportunities.
- Any *unintended consequences or indirect costs* should be considered including an outline of how these will be addressed.
- An analysis of the *enforcement arrangements* for securing compliance with the proposal, including who will enforce compliance.
- The details and results of all *consultation* exercises undertaken, including how these may have changed assumptions, costings, benefits and recommendations.
- The *recommended preferred option*, giving reasons why this was chosen.
- Details of how the preferred option will be *monitored* and *evaluated*, especially the *feedback mechanisms* to policy makers that will be established so that adjustments to the policy can be made if necessary.
- The signature of the responsible Ministers, establishing ownership of the policy and the RIA.

*Advantages and Shortcomings of Regulatory Impact Assessments*

This wide ranging set of requirements provides a structured way for policy makers to determine the potential additional burdens of policies, programmes and projects. It also has the advantage of getting policy makers to think through the full implications and impacts of

policy initiatives in a way that is similar to what the evaluation community calls a theory of change, or programme theory, approach (Chen, 1990; Weiss, 2000; Rogers *et al*, 2000). The breadth of issues that RIAs require to be taken into account is considerable and, if undertaken fully, can themselves constitute a considerable burden on policy makers and analysts. Consequently, it may be the case that not all RIAs are undertaken with the thoroughness that is expected, and/or that they might be used to justify a preferred option without the detailed consideration of all possible options that is required by the RIA process. RIAs might also lead to a 'tick box' culture in which people comply with most of the requirements of the RIA but in a superficial way and without the detailed consideration, appraisal and evaluation that is expected.

A recent evaluation of RIAs by the UK National Audit Office (2004) noted considerable variation in the quality and thoroughness of the RIA process amongst the ten RIAs that it reviewed. The NAO report noted that "departments could gain most from the RIA process if it was properly planned and resourced, and started early enough to form a genuine part of the decision making process" (NAO, 2004:4). It also noted that only two of the ten RIAs reviewed considered the range of options expected on an RIA, and that in the remaining cases only the relevant department's preferred option was assessed. There was little consideration of the 'do nothing' option. The NAO Report found RIAs to have unclear objectives, vague risk assessments, poor consideration of the counterfactual, lack of quantified assessments, poor quality cost-benefit analysis (especially analyses of benefits), and inadequate consideration of enforcement arrangements and sanctions. The details of how the impact of policies was to be monitored and evaluated were also considered brief and vague. More positively, the NAO report found that consultation was consistently the strongest element of the RIA process. Also, it was able to conclude that civil servants should "continue to see RIAs as an important part of the regulatory process", and that a "culture of scrutinising regulatory proposals" should be encouraged (*ibid*).

A separate report on Regulatory Impact Assessments by the British Chambers of Commerce (Ambler, Chittenden and Obodovski, 2004), also noted some weaknesses with the RIA process. These included poor quantification of the benefits of the regulations proposed, and a

considerable rise in the costs of regulations. The BCC Report also found that "some departments were either under-resourced or badly managed so far as preparing and archiving RIAs is concerned", and that "in many cases completion of RIAs remains a bureaucratic task to be dispatched with as little effort as possible" (Ambler, Chittenden and Obodovski, 2004:3). Like the NAO Report, the BCC review of RIAs noted that "the RIA system is an admirable means of pre-legislative scrutiny and should be extended EU-wide", whilst at the same time suggesting that "there is plenty of room for improvement in practice" (*ibid*).

**Implementation Evaluation**

The UK Government, like most other governments, is not only interested in the likely and actual impact of policies, programmes and projects; it is also interested in establishing how to successfully implement such initiatives. That is, it needs to know how, why, with whom, and under what conditions, a policy or programme can be successfully implemented and delivered. The importance of effective implementation and delivery has been highlighted in the U.K. since the General Election of 2001, when the reform and delivery of public services became the defining theme of the second Blair administration. A recent review of the evidence on effective implementation, however, has described the field as 'imperfect' and often inconclusive (Grimshaw *et al*, 2003). There is a very strong need for more and better implementation studies that can identify the particular conditions under which successful implementation and delivery takes place, or fails to take place, as well as those conditions that are more generalisable.

Successful implementation and successful impacts are clearly closely linked, and both require a combination of experimental/quasi-experimental methods of evaluation and more qualitative methods. Experimental and quasi-experimental research designs can greatly help implementation and delivery issues by bringing a degree of comparative rigour to different modes of practice. High quality qualitative data are also required using in-depth interviews, focus groups, other consultative methods (such as the Delphi and Nominal Group methods), observational methods, participant-observation methods, and social surveys.

Many, if not most, UK Government evaluations use qualitative evaluation methods to help identify effective and ineffective means of implementing and delivering policy, and to provide richer and deeper data on the impact of policies, programmes and projects. The UK Sure Start Programme, for instance, is undertaking a national impact evaluation using a quasi-experimental design, but is also undertaking in-depth case studies, local context evaluations and thematic evaluations using in-depth interviewers with providers and users of Sure Start initiatives. These more qualitative evaluations also use participant observational techniques and ethnographic methods to gather more detailed data on how Sure Start initiatives have worked (or failed to work), and to identify the conditions that are necessary to make them work elsewhere.

The UK New Deal for Communities policy, and a number of policies, programmes and projects being initiated by the UK Neighbourhood Renewal Unit), have used qualitative evaluation methods and action research to identify the delivery mechanisms and impacts of these initiatives. They have also been used to capture the local knowledge and tacit knowledge of community workers who work in local strategic partnerships (LSPs) to deliver neighbourhood renewal initiatives and to rebuild community capacity. Qualitative methods are also used extensively by UK Government evaluators to evaluate policies, programmes and projects in education, crime and justice, social welfare, health and health care, transport and environment.

The UK Cabinet Office has published a framework for using qualitative research and evaluation (Spencer *et al*, 2003) to ensure that this type of work is undertaken to agreed high quality standards. It is hoped that this and other developments, such as work on meta-ethnography (Britton *et al*, 2002; Campbell *et al*, 2003) and on including qualitative data in systematic reviews (Dixon-Woods, 2001; Harden *et al*, 2003), will enhance primary research and research synthesis that uses qualitative methods.

**Performance Managed Government**

A major feature of current UK Government policy links the allocation of resources to the performance of departments, agencies and service delivery units. These administrative and

delivery units are set performance targets that are evaluated regularly using both impact and implementation evaluation methods. Successful delivery of effective and efficient programmes or services (i.e. those that meet or exceed their targets) are rewarded with financial resources in future spending rounds. Services and programmes that do not meet their targets are either discontinued (i.e. resources are no longer allocated to them), or they are evaluated further to identify *why* the targets have not been met.

*The Policy Context – Public Spending and Fiscal Frameworks*

The United Kingdom Government has used performance management since May 1997 when the Labour Government of Tony Blair was elected into office. This approach has to be seen within the context of the UK Government's strategy for public spending and taxation. The UK public spending and fiscal frameworks seek to improve the quality and cost-effectiveness of public services while ensuring sound public finances. There are four principles underlying the public spending framework (HM Treasury, 2003). First, to provide a long-term and transparent regime for managing the public finances. Second, to measure success in terms of policy *outcomes* rather than resource inputs. Third, to provide strong incentives for government departments and service delivery units to plan over several years (up to five years or more), and to plan together where necessary. Fourth, to properly cost and manage capital assets so as to provide the right incentives for public investment.

Alongside this public spending framework is a fiscal framework that has two policy objectives. In the medium term the fiscal objective is to ensure sound public finances and that spending and taxation impact fairly within and between generations. In the short term, fiscal policy supports monetary policy and allows the automatic economic stabilisers to help smooth the path of the economy. These two objectives are implemented through two fiscal rules; the *golden rule* and the *sustainable development rule*. The former requires that over the economic cycle the Government will borrow only to invest and not to fund current spending. The latter requires that public sector net debt as a proportion of GDP will be held over the economic cycle at a stable and 'prudent' level (i.e. below 40 per cent of GDP).

*Public Service Agreements and Service Delivery Agreements*

In order to meet these public spending and fiscal objectives the UK Government has set up a performance management system around Public Service Agreements (PSAs) and Service Delivery Agreements (SDAs). PSAs set out the Government's key priorities and provide the public with a clear indication of what it can expect the Government to deliver with the resources that are available. Each large Government Department has a PSA that sets out an aim, a number of objectives, and up to ten performance targets that are linked to the PSA objectives and are *outcome* focused. In addition, each PSA has a value for money (VfM) target that establishes the cost-effectiveness of policy initiatives and the services that are delivered. This has to be done in accordance with Treasury guidance on policy appraisal and evaluation, which is published in a document known as *The Green Book* (HM Treasury, 2003a). PSAs also include a statement of who is responsible for the delivery of the targets (usually the Minister responsible for the department or policy area). In addition to departmental PSAs there are cross-cutting PSAs where responsibility for delivery is shared between two or more Ministers. Currently, there are four cross-cutting PSAs covering childcare and early years, crime and justice, action against illegal drugs, and local government.

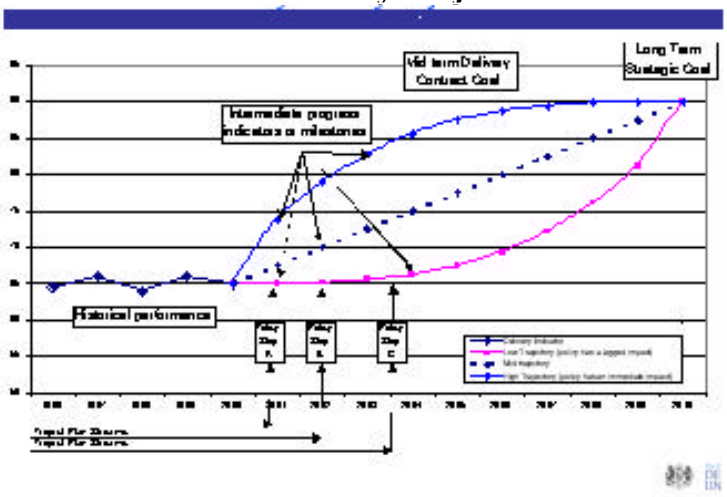
Service Delivery Agreements (SDAs) set out key programmes or actions that government departments will undertake in order to deliver their high level PSA targets. For smaller departments that do not have PSA targets, the SDA usually sets out the organisation's key targets and objectives. SDAs, like PSAs, are agreed with the UK Treasury along with measures to monitor and evaluate their progress. Every two years, as part of the Spending Review process, departments are required to establish baseline information on their work programmes and on how they propose to spend their financial allocations. Evidence must also be provided to support these plans. Departments must also provide proposals for improving efficiency as well as details of their administration budgets, investment plans and details of capital expenditure. The biennial Spending Reviews assess whether these baseline plans have been met and whether the PSA and SDA targets have been achieved. Public spending by the government is set and allocated according to success or failure of these plans and targets, and the evidence that is provided by departments. This evidence comes from national statistics,

academic research, economic theory, pilots, evaluations of past policies, commissioned research and systematic interviews with delivery agents.

*Monitoring and Evaluating Performance Management*

The monitoring and evaluation of PSA targets on a more regular basis than biennially is undertaken by the Prime Minister's Delivery Unit (PMDU), which is part of the UK Cabinet Office and works in close partnership with the Treasury (indeed it is housed at Treasury). The principal objective of the PMDU is to "improve public services by working with departments to help them meet their PSA targets, consistently with the fiscal rules" (PMDU, 2003). It does this by conducting priority reviews, stocktakes of progress towards achieving targets, and challenge meetings with departments and delivery agents. Departments are required to produce trajectories (Figure 6) showing the anticipated progress towards a target or long-term strategic goal, against which actual progress can be mapped. These trajectories have to be based on historical performance. Good quality, timely data have to be collected by departments to track progress against trajectories. Regular monitoring and reviews of these trajectories allow for prompt diagnosis and problem solving of failures in progress towards delivery targets. Continual problems with progress require more detailed analysis and plans for remedial action,

**Figure 6**
**Public Delivery Trajectories**



The performance management system used by the UK Government uses a range of evaluation methods. In terms of *impact* evaluation the target setting and target monitoring that is so central to the performance management process is essentially what evaluation specialists call goals-based evaluation, or 'legislative monitoring' (Patton, 2002). The value for money part of the performance management system uses cost-effectiveness analysis and other tools of economic evaluation. These impact assessments are backed up by evaluations of the effectiveness of interventions that have used experimental and quasi-experimental methods as well as systematic reviews and meta-analyses of existing evidence. Statistical and econometric modelling are also used to evaluate the likely and actual impacts of policies under different assumptions and tests of sensitivity.

In terms of *formative*, or *process*, evaluation, the UK performance management system uses in-depth interviews and focus groups with key stakeholders in the delivery process, including strategic planners, policy makers, delivery agents and front life service staff. It also uses documentary analysis, observational analysis (site visits), and analysis of administrative and survey data.

*Strengths and Weaknesses of Performance Management*

Performance management provides a relatively clear, transparent and outcomes-focused approach to evaluation for government. Its shortcomings include a reductionist mentality, goal or policy displacement, perverse incentives, failure to identify unanticipated outcomes, and potential insensitivity to the problems of delivery and the underlying causes of delivery failure. Target setting can become an end in itself and reduce the complexity of policy development and delivery to a few arbitrary achievements. This can result in goal or policy displacement whereby the pursuit of a given goal or target leads to resources and concentration of effort being withdrawn from other laudable goals in order to meet the targets that are being measured. This, in turn, can lead to unanticipated consequences such as harming the delivery and outcomes of services to people with these other conditions, needs and wants. Moreover, if the targets that are set are insensitive to the needs, values and demands of the users of public services, they can result in the charge that the government has "hit the target but missed the point".

The U.K. Government is aware of these shortcomings of performance management, and has taken steps to avoid some of them. Treasury guidance on the setting of PSA targets, for instance, suggests that they should be SMART (Specific, Measurable, Achievable, Relevant, and Timed), and that they should "not be open to distortion" (HM Treasury, 2003b). In terms of the latter, this means that PSAs:

> "should not create perverse incentives or encourage staff to massage or misrepresent performance data; encourage staff to focus on easy-win cases above more problematic and important cases; or lead people to compromise quality in order to achieve a measured target" (*op cit*).

Nonetheless, the UK National Audit Office (2001) found that "departments faced challenges in devising measures which are shared or influenced by other departments, which capture the essence of their objectives and which can be implemented in ways which avoid promoting perverse behaviours." The same report also noted, however, that "some changes have already taken place" and that "by refining the application of outcome-focused targets, drawing on the emerging good practices identified in this report and elsewhere, there is the prospect of more firmly evidenced improvement in performance in future" (*op cit:* 9).

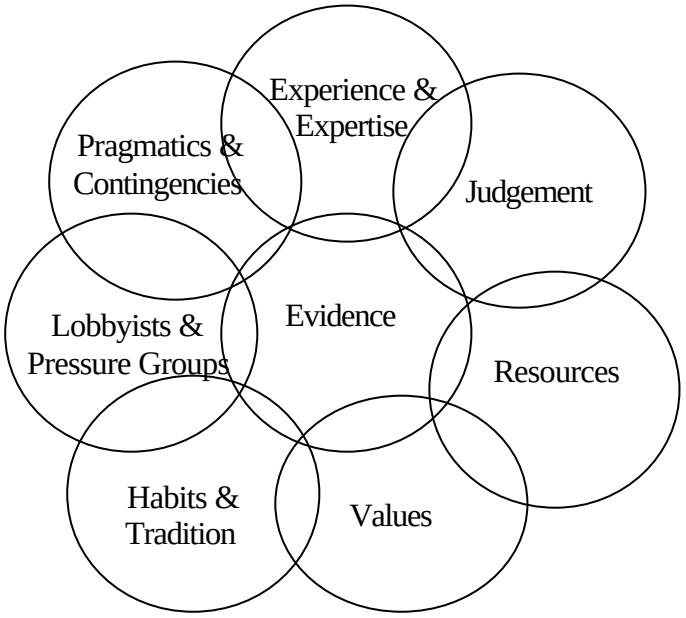
**Factors Other Than Evaluation**

Evaluation and evidence are not the only factors that influence policy making and service delivery in the U.K[1]. The experience, expertise and judgement of policy makers, and those people who have responsibility for planning and delivering policies and public services, are important factors in the policy making process. So too are the finite resources that are available for policies, programmes and projects. The values and value system within which contemporary politics take place are also contributory factors to the policy making process. This includes beliefs, ideologies, and party manifesto commitments. Policy making also involves habitual and traditional ways of doing things that may sometimes defy rational explanation yet nonetheless exist and often define what can and cannot be done in making and implementing policy. The influence of lobbyists and pressure groups on policy making cannot

[1] A fuller discussion of the other factors that influence policy making is provided in Davies (2004).

be under-estimated and this can bring selective and unevaluated evidence to the policy making process. Finally, the policy making process can be strongly affected by unforeseen circumstances and contingencies, the response to which can sometimes be opportunistic rather than well thought through, soundly evaluated, and evidence based. A graphical representation of the factors influencing policy making in the U.K. (and probably many other countries) is presented in Figure 7. Failure to appreciate the importance of these other factors that influence policy making can result in an over-estimation of role of evaluation and evidence in the policy making process.

**Figure 7**
**Factors Influencing Policy Evaluation in the United Kingdom**




Source: Davies, P.T. 2004

## Conclusion

The UK Government has a well developed system of policy evaluation, the principal aim of which is to provide effective and efficient policies, programmes and public services. Policy evaluation in the UK is undertaken to support both strategic development and operational management of public policy and public services. The full range of evaluation methods is used by UK policy evaluators, including experimental and quasi-experimental designs, social surveys, qualitative methods, economic evaluation methods, benchmarking, regulatory impact assessments and performance management procedures.

Policy evaluation in the UK is undertaken by both internal and external analysts and is increasingly subject to quality control procedures operated by the Cabinet Office (Government Chief Social Researcher's Office), HM Treasury (Government Chief Economist's Office), and the Office of National Statistics (the National Statistician's Office). Quality control is organised by providing up-to-date guidance on evaluation methods, professional development courses for government analysts, and advice and consultation from the Government Chief Social Researcher's Office and other central agencies.

Despite this well developed system of policy evaluation in the UK it would be naïve to assume that it is the only contributory factor to policy making and service delivery in the UK. Policy evaluation and evidence-based policy making has to compete with other factors, such as the experience, judgement and habitual methods of policy makers and people who deliver policies and services. The values, beliefs and ideologies that define the political environment within which policy making takes place are also important factors with which policy evaluation has to compete. Failure to appreciate the power and influence of these other factors will give a distorted view of the role of policy evaluation in UK Government.

**References**

Ambler, T., Chittenden, F., and Obodovski, M., 2004
*Are The Regulators Raising Their Game: UK Regulatory Impact Assessments in 2002/3,* Report by the British Chambers of Commerce, London, British Chambers of Commerce.

Britten N, Campbell R, Pope C, Donovan J, Morgan M, Pill R., 2003.
'Synthesis of qualitative research: a worked example using meta ethnography', Journal of Health Services Research and Policy, **56**, 4, 671-684.

Cabinet Office 1999a
*Modernising Government,* White Paper, London, Cabinet Office.

Cabinet Office, 1999b
*Professional Policy Making for the Twenty-First Century,* Cabinet Office Strategic Policy Making Team, London, Cabinet Office.

Cabinet Office, 2000
*Adding It Up,* London, Performance and Innovation Unit, Cabinet Office.

Cabinet Office, 2001b
*Better Policy Making,* Centre for Management and Policy Studies, London, Cabinet Office.

Cabinet Office, 2003a
The Magenta Book: A Guide to Policy Evaluation, Cabinet Office, London, Government Chief Social Researcher's Office, Available at http://www.policyhub.gov.uk.

Cabinet Office, 2003b
*Trying It Out: The Role of 'Pilots' in Policy Making*, London, Cabinet Office, Government Chief Social Researcher's Office.

Cabinet Office, 2003c
*Strategic Audit, Discussion Document*, London, Cabinet Office, Prime Minister's Strategy Unit.

Cabinet Office, 2003d,
*Better Policy Making: A Guide to Regulatory Impact Assessment*, London, Cabinet Office, Regulatory Impact Unit.

Cabinet Office, 2004
*Strategy Survival Guide,* London, Cabinet Office, Prime Minister's Strategy Unit.

Campbell R, Pound P, Pope C, Britten N, Pill R, Morgan M, Donovan J., 2002
Evaluating meta-ethnography: a synthesis of qualitative research on lay experiences of diabetes and diabetes care. Social Science and Medicine, **7**, 4, 209-215.

Chen, H.T., 1990
Theory-Driven Evaluations, Newbury Park, Sage Publications.

Davies, P.T., 2004
*Is Evidence-Based Government Possible?* Jerry Lee Lecture, Presented at the 4th Annual Campbell Collaboration Colloquium, Washington D.C., February 18-20, 2004.

DfES, 2002
*Education Maintenance Allowance: The First Two Years A Quantitative Evaluation*, Research Report RR352, London, Department for Education and Skills.

Dixon-Woods M, Fitzpatrick R, Roberts K., 2001
'Including Qualitative Research In Systematic Reviews: Problems and Opportunities'. Journal of Evaluation in Clinical Practice, **7**, 125-133

Grimshaw, J.M., Thomas, R.E., MacLennan, G., Fraser, C., and Ramsay, C.R., 2003
'Effectiveness and Efficiency of Guideline Dissemination and Implementation Strategies', Final Report, Aberdeen, Health Services Research Unit.

Harden, A., Oliver, S., Rees, R, Shepherd, J., Brunton, G., Garcia, J., and Oakley, A., 2003
'An Emerging Framework for Synthesising the Findings of Different Types of Research in Systematic Reviews for Public Policy', Paper presented at the 3rd Annual Campbell Colloquium, Stockholm, Sweden (available at http://campbellcollaboration.org).

HM Treasury, 2003a
*The Green Book: A Guide to Appraisal and Evaluation*, London, HM Treasury.

HM Treasury, 2003b
*Public Spending Guidance*, London, HM Treasury (www.hm-treasury.gov.uk).

HM Treasury, 2004
*Public Spending Guidance*, London, HM Tresury.

Morris, S., Greenberg, D., Riccio, J., Mittra, B., Green, H., Lissenburg, S., and Blundell, R., 2004
*Designing a Demonstration Project: An Employment, Retention and Advancement Demonstration for Great Britain*, London, Cabinet Office, Government Chief Social Researcher's Office, Occasional Paper No. 1 (2nd Edition).

National Audit Office, 2001
*Measuring the Performance of Government Departments*, Report by the Comptroller and Auditor General, HC301, Session 2000-2001, 22 March 2001, London, National Audit Office.

National Audit Office, 2004
*Evaluation of Regulatory Impact Assessments Compendium Report 2003-04,* HC 358, Report by the Comptroller and Auditor General, HC301, 4 March 2004, London, National Audit Office.

Stephenson, et al. 2004
A school-based randomised controlled trial of peer-led sex education in England. *Controlled Clinical Trials*. (In Press)

Patton, M.Q. 2002
*Qualitative Research and Evaluation Methods*, 3rd Edition, Thousand Oaks, Sage.

PMDU, 2003
*Monitoring, Analysis, Evaluation & Reporting***,** PMDU Toolbox, London, Prime Minister's Delivery Unit.

Rogers, P.T., Hacsi, T.A., Petrosino, A., and Huebner, T.A. (eds), 2000,
'Program Theory in Evaluation: Challenges and Opportunities', New Directions for Evaluation, No 87, San Francisco, Jossey-Bass.

Spencer, L., Ritchie, J., Lewis, J., and Dillon, L., 2003
Quality in Qualitative Evaluation: A Framework for Asssessing Research Evidence',
Government Chief Social Researcher's Office, London, Cabinet Office

Rosenbaum, P. and Rubin, D.B., 1985
'Constructing a Control Group Using Multivariate Matched Sampling Methods that Incorporate the Propensity Score', *American Statisti*cian, 39-38.

Shadish, W.R., Cook, T.D., and Campbell, D.T. 2002
*Experimental and Quasi-Experimental Designs for Generelised Causal Inference*, Boston, Houghton Mifflin Company.

Weiss, C.H. 2000
'Which Links in Which Theories Shall We Evaluate?', in Patricia J. Rogers, Timothy A. Hacsi, Anthony Petrosino, Tracy A. Huebner (eds*), Program Theory in Evaluation Challenges and Opportunities*: New Directions for Evaluation, No. 87, Boston, Jossey-Bass.