Work in progress. Not for quotation or citation.
All comments are welcome: h.s.t.mcleod@bham.ac.uk

Evaluating the outcomes of policy initiatives: lessons for modernising the NHS

Hugh McLeod
Health Economics Facility, HSMC

Introduction

Policy initiatives to improve the delivery of health care in the NHS have accumulated considerable momentum over the last two decades. Whether couched in terms of quality, patient choice, efficiency or cost-effectiveness, policy makers have attempted to bring about change in the behaviour of clinical and non-clinical staff via a range of mechanisms. One set of mechanisms relates to structurally-orientated change (Griffiths' general management, the purchaser/provider split, primary care trusts). Another set relates to process-orientated change (often using 'quality improvement' tools for influencing working practices based on techniques developed to increase competitiveness in non-health care industries). The process-orientated mechanisms are increasingly combined with a third set of mechanisms relating to national targets and performance measures.

This paper first outlines how the policy-making process and the role of evaluation has evolved during this period, and then describes Pawson and Tilley's (1997) approach to 'realistic' evaluation. The paper then describes several major evaluations and their conclusions in the light of Pawson and Tilley's conceptual framework for 'realistic evaluation'. On the basis of these evaluations and others of recent policy programmes lessons relating to the conduct of pilot programmes and evaluations are then outlined.

The policy-making process

After the 1974 reorganisation, which exemplifies the 'top-down' model of policy making in which policy formation is distinct from subsequent implementation[1], the style of policy making took "the form of a series of 'policy accidents' as a result of which new policies emerged in a somewhat unplanned fashion" (Harrison and Wood (1999, p755). For example, Griffiths' policy recommendations in 1983 to introduce individual general managers in place of the system of consensus team decision making were "radical", but also "vague … and contained only the sketchiest account of the functions of various new institutions" (Harrison and Wood, 1999, p756). Policies such as this

> "were necessarily defined by implementation at least as much as by their original content and they therefore represent a weakening of the policy/action dichotomy which had characterized the earlier period. It is important to note that … the dissolution seems to have occurred accidentally rather than as the article of policy which it later became" (Harrison and Wood, 1999, pp755-6).

Enthoven (1985, p21) argued that Griffiths' management proposals did not go far enough[2], and suggested a key role "for 'market forces' or some motivating factor that

[1] The April 1974 reorganisation was "literally conducted according to a blueprint which had been carefully developed over a period of several years" (Harrison and Wood, 1999, p752).

[2] "If the structures and incentives in the NHS are not changed more fundamentally, these recommendations are not likely to change much. There is a real danger they will be little more than cosmetic" (Enthoven, 1985, p19).

serves the same purpose". Implementation of the 1991 'internal market'[3] was staggered on the basis of volunteer waves of both hospital trusts and general practitioner (GP) fundholding. This innovation allowed for policy to evolve in the light of the pioneers' experience. Harrison and Wood gave the name 'manipulated emergence' to this style of policy making in which

> "the expert contribution lay in the translation by incentivised local actors of the bright idea into specific organizational arrangements which accord with the philosophy behind the original idea" (1999, p752).

The use of volunteers in waves to pilot new initiatives has been widely employed since, by both Conservative and Labour governments, although the extent of subsequent rollout has varied. While the introduction of Primary Care Groups (PCGs) in April 1999 was obligatory, the hierarchy of four levels of responsibility provided an opportunity for individual PCGs to develop at different rates, on the basis of voluntary action. In practice, however, there is evidence of pressure to be seen to be making rapid progress (Regen, 2001).

The 'manipulated emergence' approach to policy making has also been applied to major process-orientated initiatives, which, under the auspices of the National Patient's Access Team (NPAT) (and subsequently the NHS Modernisation Agency (MA)), have tended to be rolled out rapidly (albeit very thinly). For example, the National Booked Admissions Programme (NBAP) included some booking in all trusts in England by the third wave, and the Cancer Services Collaborative (CSC) achieved partial coverage of all cancer networks in two phases.

The role of evaluation

Thatcher's personal support for Griffiths' management reform was decisive in the face of "a good deal of resistance from the NHS over a prolonged period, and despite the diffidence of the secretary of state" (Harrison and Wood, 1999, p756)[4]. Thatcher's 'politics of conviction' was not inconsistent with Griffiths' views on research evidence: he was "opposed on principle to pilot studies on the ground that they served merely to obstruct or delay implementation" (Harrison and Wood, 1999, p763)[5].

In contrast, Enthoven called for demonstration projects as a basis for testing policy initiatives, "not only to see *whether* they work, but to determine exactly *how* to make them work" (1985, p28).

> "Better to begin with a dozen pilot Districts whose managements are enthusiastic about the idea, develop and test it with the benefit of expert advice, then push it to the maximum in the pilot districts, and display the benefits for all to see" (Enthoven, 1985, p2).[6]

---

[3] *Working for Patients* (Secretaries of State for Health, 1989) "was brief and vague, the very antithesis of a blueprint, … the fragmented product of policy making on the hoof" (Harrison and Wood, 1999, p752).

[4] Enthoven was also against the rapid national implementation of Griffiths' plan: "And why must it be achieved by the end of 1985, if other than for political effect?" (1985, p20).

[5] Griffiths viewed analysis and enquiry as important, but that they "should be finite and that action should follow; [as Griffiths put it] 'if too much time is allowed for expression, people sabotage things, and anyway, research is rarely conclusive….'" (Wistow and Harrison, 1998, p665).

[6] Enthoven (1998, p20) " A more effective way to implement the idea would have been to work with a few interested pilot Districts or perhaps one Region, and work out in detail exactly how it should be done including job descriptions, delegations of authority, and reporting relationships. Then try it for a

While Enthoven's policy ideas provided a basis for the 1991 'internal market', his views on evaluation were not accepted in the face of the Prime Minister's and Secretary of State's preferences.[7] Nevertheless, evaluation of pilot initiatives became increasingly common. For example, the Department of Health (DH) commissioned an evaluation of the total quality management (TQM) initiatives in 38 hospitals between 1990 and 1993 (Joss and Kogan, 1995)[8]. Similarly, in 1993, the DH agreed to support the piloting and evaluation of business process re-engineering[9] (BPR) at both the Leicester Royal Infirmary (Bowns and McNulty, 1999) and King's Healthcare Trusts (Packwood et al, 1998). In October 1994, the NHS Executive (1994) announced its intention to pilot and evaluate the 'total' purchasing pilots (TPPs), as one option for developing the purchaser/provider split. Evaluation was also promised for the PMS, PDS and GP locality pilots[10].

The increasing government commitment to both piloting and evaluating new initiatives coincided with increasing calls for 'evidence-based policy' (Ham et al, 1995). While Black (2001) and others have set out the challenges associated with developing evidence-based policy, the Labour government has made its objectives explicit:

> "Government should regard policy making as a continuous, learning process, not as a series of one-off initiatives. We will improve our use of evidence and research so that we understand better the problems we are trying to address. We must make more use of pilot schemes to encourage innovations and test whether they work. We will ensure that all policies and programmes are clearly specified and evaluated, and the lessons of success and failure are communicated and acted upon" (Prime Minister and the Minister for the Cabinet Office, 1999, p17).

The evaluation of major pilot programmes focusing on process-orientated initiatives, including the first wave of the NBAP (Ham et al, 2002) and phase I of the CSC (Robert et al, forthcoming), can be viewed in this context.

Pawson and Tilley's 'realistic evaluation'

Pawson and Tilley (1997) set out a conceptual framework for the evaluation of complex social programmes. They argue that the outcomes of a programme can only be explained in terms of the interplay between mechanisms and contexts:

> "Evaluators will always construct their explanations around the three crucial ingredients of any initiative: context (C), mechanism (M) and outcome (O). There will always be contextual variation within and between programs, a corresponding variation in the

---

couple of years, refine the process and then, if the idea still looks good, recommend it for wider adoption."

[7] Griffiths also maintained a key role, being the only non-politician member of the Prime Ministerial Review of the NHS.

[8] Joss and Kogan (1995, p69) suggest that the main reason why the DH "chose TQM as a mechanism for generating change in the NHS …. was that key features of government reform in the NHS resonated strongly with some of the requirements of TQM".

[9] BPR was defined by its authors (Hammer and Champy, 1993, p32) as "the fundamental rethinking and radical redesign of business processes to achieve dramatic improvements in critical, contemporary measures of performance, such as cost, quality, service, and speed".

[10] By November 1995, Malone (1995) had confirmed the government's intention to evaluate the PDS pilots (which were subject to new legislation), as was also the case with the PMS pilots. In December 1996, the Secretary of State announced that GP locality pilots (subsequently labelled GP commissioning groups) would be created and evaluated (Secretary of State for Health, 1996). By the introduction of PCGs in 1999, the presence of an evaluation was inevitable (not least because the evaluation of GP commissioning groups still had a year to run).

> effectiveness of causal mechanisms triggered, and a consequential variation in patterns of outcomes, giving realist research the task of modelling the different ways in which the Ms, Cs and Os come together. … [W]e term those propositions which combine this trio of explanatory components context-mechanism-outcome pattern configurations or CMO configurations. … The task of a realist evaluation is to find ways of identifying, articulating, testing and refining conjectured CMO configurations" (Pawson and Tilley, 1997, p77).

Pawson and Tilly (1997, p107) argue that as evaluators generate their initial hypotheses about CMO configurations which provide the most compelling possibilities for change, they should draw on both academic literature and "the 'folk wisdom' of practitioners". Pawson and Tilly illustrate their ideas using their experience of evaluating programmes relating to the criminal justice system. For example, an evaluation of the impact on the recidivism of prison inmates who participated in a long standing higher education programme, started

> "with a period of theory development, derived from qualitative investigation, in which educators were interviewed to elicit realistic theories on program mechanisms ('what it was about the programme which might generate change in a prisoner') and program contexts ('with which sorts of inmates in which conditions the initiative might be successful'). These 'folk theories' were then used to interrogate a range of data in a range of ways" (Pawson and Tilly, 1997, p88).

This activity is based on relevant theory, such as, for example, utility maximisation: "the venerable analytic frameworks which have tried to explain how we humans make choices" (Pawson and Tilly, 1997, p122). This allows the evaluator to draw on "middle-range theory" such that "from a small core of ideas, it is possible to develop a wide range of testable propositions. In realistic evaluation terms, this means that we build up families of [CMO] configurations" (Pawson and Tilly, 1997, p123).

> "Each evaluation within a problem area is seen as a case study, and the function of the case is to refine our understanding of the range of CMOs which seem to have application in that domain. We derive, quite literally, a catalogue of answers to the question of what works for whom in what circumstances … amassed in a manner which involves ascending and descending the route between abstraction and specification" (Pawson and Tilly, 1997, p125).

This focus on addressing the question of 'what works for whom in what circumstances' renders the task of outcome analysis one of building up a picture of the key causal mechanism and context configurations associated with different outcomes:

> "The key point of design … is that the subgroup analysis is able to demonstrate *prodigious difference in levels of success achieved within the program*. … [W]ithin all programs, there will be many, many different groups of winners and losers. Identifying these subgroups for whom the program succeeds (or indeed fails) gives us a window on why the initiative works. The research strategy works by dint of some prior theoretic work which enables researchers to get an approximate fix on the likely successes and failures. Sometimes the relative performances will confirm expectations, sometimes they will lead to a readjustment of the 'folk theories'. … Outcomes only follow when particular mechanisms have been triggered in particular contexts, and they will only reveal themselves when investigation has traced them through the same pathway" (Pawson and Tilly, 1997, pp113-4).

This position is set out in contrast to one in which programmes are viewed "as some kind of unitary happening which either does or doesn't work" (Pawson and Tilly, 1997, p104).

The goal is to identify regularities in the relationships between context, mechanism and outcome within complex social programmes, in order to inform the choice of mechanisms for a given context to achieve desired outcomes.

Pawson and Tilly (1997, p85) see themselves as "whole-heartedly pluralists when it comes to the choice of method" for testing hypotheses about CMO configurations, in order to ensure that the method is "carefully tailored to the exact form of hypothesis". Similarly, data collection is focused on the objectives of testing and refining CMO configurations.

Pawson and Tilly remain acutely aware of the potential for evaluation findings to be unfulfilled in the world of 'realpolitik'.

'Realistic evaluation' of health care programmes

Several evaluations of major health care programmes have explicitly attempted to explain outcomes in terms of the interplay of mechanisms and contexts. The evaluations of the first waves of TPPs (Mays et al, 2001b) and the NBAP (Ham et al, 2002) are discussed below. Judge (2000) reported the intention to use the CMO framework to inform the national evaluation of Health Action Zones. The final report of this evaluation is due to be submitted to the DH in December 2002. It is also apparent that Pawson and Tilly's approach is influencing the specification of evaluation research by the DH[11].

The 'total' purchasing pilots

In response to several influential local initiatives[12] the NHS Executive issued a call for fundholding practices, either singly or in groups, to volunteer to purchase potentially all the hospital and community health services for the patients on their practice lists. The resulting two waves of TPPs, 1995/96 to 1997/98, were the subject of a major DH-funded evaluation; the first of the 'internal market' developments to be so treated.

As Mays et al (2001a, p417) point out, in commissioning the TPP evaluation

> "the motives of the policy makers were unclear. They may have genuinely wished an honest answer to the question, 'Is total purchasing more cost-effective than the status quo?'. An alternative, and equally plausible, view is that by commissioning the evaluation they hoped to disarm their political opponents and critics of the existing fundholding scheme who were crying out for evaluation at the time. While being free to implement their desired policy, the government could be secure in the knowledge that the evaluation would not report until long after the in policy had been fully implemented!"

While the government's decision not to evaluate GP fundholding was repeatedly made an issue by opposition spokesmen[13], the TP evaluation did directly inform policy, but that of the Labour government before the publication of the *New NHS* white paper in 1997 (Mays et al, 2001a).

---

[11] See, for example, the Research Specification for the evaluation of the Changing Workforce Programme (Department of Health, 2002c).

[12] By early 1993, several groups of fundholding GPs wanted to extend their influence over the use of NHS resources and demonstrate that general practitioner-led purchasing and service development could improve a wider range of services (Bosanquet et al, 1996; Walsh et al, 1999). They succeeded in agreeing devolved budgetary responsibility for a range of emergency or unplanned hospital and community health services with their local health authorities.

[13] For example, Chris Smith debate on the Queen's Speech Hansard 25 October 1996 : Column 247 www.parliament.the-stationery-office.co.uk/pa/cm199697/cmhansrd/vo961025/debtext/61025-04.htm

Pawson and Tilly (1997) was published during the evaluation by the total purchasing national evaluation team (TP-NET):

> "In retrospect, the TP-NET evaluation would have greatly benefited from the explicit application of Pawson and Tilley's CMO framework from the outset. It would have obviated the need for a great deal of confused discussion within TP-NET and would have provided common ground for the different disciplines to share, since Pawson and Tilley's approach allows for the explicit identification of a wide range of potential 'mechanisms' (e.g. including economically and socio-logically grounded insights). The team was aware of previous theoretical and case-study work on the importance of context in policy evaluation (Pettigrew et al., 1992) and had used it to generate hypotheses about likely causes of variations between pilots as part of the 'process evaluation' … . However, before the publication of Pawson and Tilley's book, the team did not have a generic conceptual system explicitly linking context with hypotheses about processes (mechanisms) and outcomes which could accommodate both quantitative and qualitative data and which was explicitly comparative" (Mays et al, 2001a, p413).

This candid acknowledgement of the challenge of coherently combining qualitative and quantitative-based disciplines highlights the perceived strength of Pawson and Tilly's conceptual framework. Furthermore, it places the CMO framework in the context of the very important contribution of Pettigrew et al (1992) (discussed below). In specific terms, Mays et al (2001a, p414) note how CMO conjectures could have been made:

> "As a detailed understanding emerged of the heterogeneity of the 53 'first wave' TPPs during the first 'live' year, the emerging analysis of important contextual factors (e.g. degree of health authority support) and mechanisms (e.g. budget holding and independent contracting) could have been used to hypothesize predicted outcomes for different TPPs ".

In fact, it became clear during the first live year that outcomes relating to TPPs' use of non-elective hospital services were greatly influenced by their ability to negotiate contracts, if only because pilots without independent contracts tended to abandon aspirations relating hospital use (McLeod and Raftery, 2000)[14]. Nevertheless, the evaluation's final report described Pawson and Tilley's conceptual framework and presented its verdict on the pilots in terms of the interplay between context and mechanisms, thereby highlighting 'advantageous characteristics' associated with the pilots:

> "Holding budgets and having independent contracts, while important prerequisites for being taken seriously in the internal market, were not sufficient for effective total purchasing. The pilots also needed determined leadership, a robust management infrastructure, good information systems, structures to enable wide GP involvement in decision making and budgetary management, a supportive health authority and well-developed relationships with other provider organizations to make headway" (Mays et al, 2001b, p296).

The first wave pilots of the national booked admissions programme

'Booking'[15] is intended to make accessing the NHS more convenient for patients and reduce patients failing to attend, cancellations and time spent waiting (Department of

[14] In general, those pilots that did negotiate contracts found that they would not release the anticipated resources, and so only pursued objectives relating to hospital services if they could negotiate resources from the local HA.

[15] Booking refers to the agreement of hospital appointment or admission dates with patients at the time the decision to refer or admit is made. This arrangement differs from the traditional waiting list system in which patients wait before being notified, often with little notice, of the appointment or admission date.

Health, 1998a; Department of Health, 1998b)[16]. Twenty four pilots were selected to take part in the first wave of the NBAP, which ran from October 1998 to March 2000, and was led by NPAT[17] (Department of Health, 1998b). The impetus for booking increased through subsequent waves of the programme (NHS Executive, 1999; 2000; 2001) and *The NHS Plan* target of having all NHS consultant services booked by the end of 2005 (Secretary of State for Health, 2000). The final report of the DH-commissioned evaluation described the prominence that the initiative gained:

> "From small beginnings, the booked admissions programme has become a central part of the government's policy of improving access and convenience within the NHS. As the first of the national initiatives led by NPAT, the programme led the way in what became a sustained and systematic attempt to redesign services from the patient's point of view. The first wave pilots that are the focus of this study were therefore trailblazers not only for the subsequent roll out of booked admissions but also for the development of the work of NPAT and eventually the NHS Modernisation Agency as a whole" (Ham et al, 2002, p29).

In contrast to the TP initiative, the purpose of the NBAP evaluation was clear:

> "the policy context in which researchers work is dynamic and at no point was there an expectation that the world would 'stand still' until the results of our assessment of the first wave pilots became available. The establishment of successive waves of booked admissions pilots and the commitment in **The NHS Plan** to achieve 100% coverage by 2005 underscored this point and added to the potential importance of the evaluation for policy and practice. The movement of booking into the mainstream also served to confirm that the main purpose of the evaluation was to assist those in government and the NHS to enable booking to work as effectively as possible rather than to determine whether booking was the right direction of travel" (Ham et al, 2002, p249).

By seeking to explain the variations in the pilots' performance in relation to the context in which booking developed and the mechanisms used, Ham et al placed the evaluation "in the 'realistic evaluation' tradition advocated by Pawson and Tilley" and highlighted the

> "overlap between this tradition and research on organisational and service innovation (Pettigrew et al, 1992), not least in the focus on receptive contexts of change in accounting for the impact of innovation" (2002, p28)[18].

The challenge of booking related to its systematic implementation rather than the feasibility of small-scale consultant-led initiatives[19]. On the basis of a literature review and the findings of visits to eight hospitals in which consultants were booking, Bensley et al (1997) highlighted the issues of waiting times, capacity, working practices, and consultants' and patients' views. Bensley et al concluded that "the

---

[16] Although booking is not new to the NHS (Yates, 1998) and has long been thought to benefit patients (Southham and Talbot, 1980), it has not become widely established despite being incorporated in guidelines (for example RCS, 1991). With one unsuccessful, but influential, exception (South Western RHA, 1991), booking initiatives have been limited to the efforts of individual surgeons, and the subject of few analyses (Southham and Talbot, 1980; Houghton and Brodribb, 1989; Simpson and McCallum, 1992; Barnes et al, 2000).

[17] The DH provided £9.9 million to the first wave pilots.

[18] Ham et al (2002, p28) continue: "In turn, there is a link between the organisational analysis literature and research on the management of change (Iles and Sutherland, 2001). We have drawn on these various strands of work, together with research into quality improvements in health care (Shortell et al, 1995), both in organising data gathering and in trying to make sense of what the evidence we have gathered is telling us about the evolution of booked admissions."

[19] Fearn and Bensley (2000) reported a survey of trusts just before NBAP commenced and this showed that a small proportion of consultants were booking all their patients in a wide range of specialties.

introduction of a totally booked admission system would require a major organisational and cultural change programme" (1997, p14).

In addition to funding from the DH which allowed for more than project management support, NPAT's role provided a distinctive range of complementary mechanisms to support the local implementation of booking. In particular, the training opportunities for project managers and the forum for sharing ideas and experience between the pilots was valued. NPAT also provided impetus to redesign processes as a consequence of considering the patient's perspective[20].

The evaluation assessed the pilots' booking performance using a proxy measure[21] at three points in time, which corresponded to a baseline, the end of the pilot phase, and one year after the end of the pilot phase[22]. The pilots varied greatly in terms of scope, size and performance (Ham et al, 2002). The progress during the pilot phase in increasing booking for day case patients was not fully sustained, and much less progress was made with inpatient booking. Only three day case pilots stand out for having achieved a high level of booking in the post-pilot phase across more than half of all day case activity.

A wide range of issues was reported to have facilitated and inhibited booking. Pilots that achieved a high level of booking tended to have a 'receptive context' characterised by previous booking experience, effective leadership by a chief executive and senior clinicians, a dedicated project manager and team, and a flexible approach to clinicians. Pilots facing capacity constraints such as a shortage of beds and theatre time, waiting times over six months, relying on complex information and communications technology, and dealing with the consequences of trust mergers were at a disadvantage in implementing booking.

Consultants' reluctance to change practices and give up the freedom to determine relative priority available under traditional waiting list arrangements was widely reported to have slowed the implementation of booking. In some cases, this included an unwillingness to plan leave sufficiently far ahead to enable booked dates to be agreed and honoured.

[20] This emphasis stemmed from the experience of NPAT staff of BPR at the Leicester Royal Infirmary. Locally instigated initiatives included those at Central Middlesex Hospital (Layton, 1998). Sir Roy Griffiths would have approved as his key interest was in viewing services from the client or customer's point of view. An "insider" commenting on Griffiths' approach to his review of community care, stated "His focus right from the start was 'I want to look at the interface between the client and the direct service which is provided. That is the focus, the rest is all superstructure …'" (Wistow and Harrison, 1998, p654).

[21] The percentage of patients waiting with a booked or 'to come in' date.

[22] The end of March in 1999, 2000 and 2001. The resulting 'snap-shot' pictures of performance were compared with the quarterly specialty-level KH07 data in the most common specialties. The evaluation requested data directly from the pilots, which included DNAs, cancellations, waiting times and numbers, and admission numbers (Kipping et al, 2000). (The evaluation intended to use the quarterly analysis of progress supplied by each pilot to NPAT as a basis for assessing outcomes, but it became clear that this would be unsatisfactory because of the absence of standardisation in reporting across the pilots.)

Those pilots which, with few exceptions, had not yet achieved their objectives, faced a range of obstacles to making further progress[23]. This weakening of the implementation mechanisms, when accompanied by a deterioration in local context (for example, in the form of pressure on waiting times) helps to explain the poorer booking outcomes in the larger more complex pilots. Most importantly, the weakening of the implementation mechanisms will have undermined the key on-going task, which in this case was the process of negotiation with consultants to fully accept booking and its consequences.

Discussion

The TPP and NBAP first waves were both high profile attempts to introduce change to aspects of longstanding health care provision. Although the programme-specific interventions were very different, the conclusions of the TPP and NBAP evaluations are strikingly similar in terms of the 'advantageous characteristics' associated with comparatively successful pilots. Both evaluations concluded that performance could be explained in terms of effective leadership, previous relevant experience, management support, participation by clinicians, and receptive local contexts.

In addition to using Pawson and Tilly's CMO framework, both evaluations underline the importance of receptive contexts[24] for change described by Pettigrew et al (1992). Through the study of strategic change processes, Pettigrew et al (1992, p9) set out a "novel framework for researching the problem of management of change in the NHS" in which "the analytical challenge is to connect up the content, contexts and processes of change over time to explain the differential achievement of change objectives".

The Modernisation Agency's wide-scale adoption of the collaborative approach to implementing process-orientated change is informed by considerable experience including BPR and NPAT's hybrid booked admissions arrangements[25]. However, the current challenge for the Modernisation Agency is considerable. As Ham et al (2002) concluded

> "the challenge is to demonstrate that it can support the NHS not only to bring about pockets of improvement (the main achievement to date) but also organisation wide and ultimately systems wide change" (Ham et al, 2002, p271).

This view that the MA has considerable further ground to cover contrasts with the modernisation agency's own assessment of its position, illustrated here in its recent guidance for the imminent Modernising Dentistry Programme pilots:

> "The Modernisation Agency (MA) has an established track record of service improvement across a wide range of NHS services. It draws on the considerable experience and expertise within the Agency on approaches, tools and techniques of health service improvement … . … The Programme will adopt the tools and approaches that have already proved to be successful in healthcare improvement in the NHS and worldwide" (NHS Modernisation Agency, 2002a, p4).

---

[23] These included project managers being reassigned or headhunted, ongoing project funding being unclear, the transfer of NPAT's support role to regional offices, and the presence of funding for wave two projects with different objectives.

[24] "The distinction developed between receptive and non-receptive contexts for change represents a novel, distinctive and fruitful concept which needs to be confirmed, modified or indeed refuted in the light of additional evidence" (Pettigrew et al, 1992, p60).

[25] To some extent this evolution mirrors the move from the 'top-down' approach towards the 'manipulated emergence' approach to promoting structure-orientated change.

Lessons relating to several aspects of pilot programme management should be considered in order to move beyond 'pockets of improvement':

1 'Breakthrough' objectives

Berwick (1996) powerfully articulated the need to focus on 'stretch goals' in order to achieve 'breakthrough' improvements in the delivery of health care[26]. Pilot programmes should aim for substantial change when required in order to achieve 'best practice'. However, in those initiatives in which tangible 'stretch goals' are specified, evaluations consistently indicate that expectations are not met by the majority of participants. This is particularly evident in the process-oriented initiatives. For example, the emphasis on 'dramatic' improvement within short time spans is a key element of the Breakthrough Collaborative approach:

> "The … model seeks to achieve unprecedented levels of improved performance in participating organisations in less than 1 year by bringing providers together to understand and drive improvement within a specific topic area" (Kilo, 1998, p2).

This position, which appears to be widely accepted in the NHS, despite there being no independent evidence of programme-wide success within this time frame, stands in marked contrast to Berwick's earlier conclusions about the likely short term impact of quality improvement initiatives. Commenting on the outcomes of the quality improvement initiatives hosted by the Harvard Community Plan[27] in the late 1980s, Berwick et al stated

> " … because these are real stories, there are no astounding results to report. How could there be in a project that lasted only eight months? … Before disappointment wins out, however, take a closer look. By different measures - the measures of new ways of thinking, new insights about work processes, new working relationships, and small, permanent breakthroughs in understanding - by these simple, modest measures of early success, the achievements of these pioneer project teams exceeded our most optimistic hopes" (1990, pxix).

Why should 'permanent breakthroughs in understanding' routinely lead to *rapid* permanent breakthroughs in performance, given the complexity of health care services (Plsek and Wilson, 2001)? The Northern New England Cardiovascular Disease Study Group (O'Connor et al, 1996; Kasper et al, 1992) combined detailed data analysis with process-orientated change initiatives, and was highly successful in reducing hospital mortality rates for coronary artery bypass grafting (CABG) surgery[28]. While cited as evidence of the collaborative approach (Plsek, 1997; Kilo, 1999), the initiative took place over a much longer time scale than that advocated by the 'Breakthrough' collaborative model adopted by the MA.

---

[26] For example, Berwick illustrated the importance of stretch goals when commenting on the target of a 30% reduction in caesarean section rates as a collaborative breakthrough goal: "A less ambitious goal might have led simply to stressing the system to achieve marginal gains. By contrast, a safe reduction of 30% or more required fundamental changes in patient preparation, anaesthesia, labour management, and delivery technique" (1996, p621).

[27] the US National Demonstration Project on Quality Improvement in Health Care.

[28] A three-year study of variability in mortality rates across the five participating organisations convinced the members of the study group that the "observed differences in outcome were real and could not be attributed solely to differences in case mix. … Juran refers to such a conclusion as a breakthrough in attitude, and notes that such breakthroughs are prerequisite for improvement" (Plsek, 1997, p87). A period of site visits, in which each surgical team observed each other team undertaking CABG surgery, was followed by promising site-specific changes in processes being tested using continuous quality improvement tools. This process took a further two years. Subsequent analysis indicated that hospital mortality rates for CABG surgery fell by 24%.

To the extent that there may be a trade-off between the 'achievable' magnitude of the change and the time required, policy makers should reconsider their approach to pilot programmes:

2 Pilot programme myopia
Pilot programmes are intended to provide receptive contexts in which incentivised volunteers attempt to use new mechanisms to trail blaze. Increasingly, whether all the participants reach the end of the trail (or the trail reaches the desired destination) appears to be less important than leaving foot-prints that can be followed. Pilots with 'advantageous characteristics' can achieve 'quick wins' which provide, with minimum delay, the good news stories which accompany programme rollout. This approach may serve the short-term policy objective (if it is to be seen to be responding to the 'modernise!' battle cry), but the price of this short-term focus is a process that is incompatible with systematic change.

Clearly, it could be (and has been) argued that the 'disappointing' outcomes of initial 'wave' pilots do not matter as long as lessons have been learned and applied to the benefit of those following closely behind in the rollout. However, this outlook implicitly assumes that the lessons learned are sufficient to ensure that the desired outcome is subsequently achieved. In general, evaluations are not tasked to assess this, and so our insight into what subsequently happens to the pilots is limited. The booked admissions evaluation was unusual in providing this opportunity, and, in this case, the majority of pilots did not move closer to the objective of 100% booking after the pilot phase ended: booking had not become embedded in routine working practice[29].

In general, it may be that the desired outcome is never achieved. For example, in the case of GP fundholding which had a comparatively long lifespan, the annual rollout to new waves of participants, accompanied by relaxation of the entry requirements and learning from evaluations (albeit non-DH funded), did not lead to widespread changes in GP behaviour[30].

3 The 'advantageous characteristics' of pilots
Comparison of CMO configurations suggests that pilots attempting to change the delivery of health care services share similar 'advantageous characteristics', even when the type of initiative being pursued differs considerably. Both the TPP and NBAP evaluations noted the role of effective leadership, management arrangements, previous relevant experience etc.

This finding underlines the importance of strengthening leadership capacity and organisational development work within the NHS.

[29] In the case of the TPPs, we know little about the sustainability of the arrangements relating to the non-elective acute hospital services put in place by a minority of the pilots: some used fundholding savings to maintain discharge co-ordinator posts during 1998/99, in the hope that they would be retained and funded by the new PCG in 1999/00. A few pilots used the opportunity to become a PMS pilot as a way of maintaining their TP initiatives. Nothing has been published about the OSC post phase I, and the only publication relating to the CSC post phase I (NHS Modernisation Agency, 2002b) did not report on phase I projects or provide a comprehensive account of phase II performance.
[30] "Most [fundholders] have mastered the considerable administrative burden, but in purchasing terms they are only maintaining the status quo" (Audit Commission, 1996, p95).

Nevertheless, as available evidence relating to outcomes is restricted to those associated with short-term pilot programmes, it is perhaps inevitable that these ‘generic’ advantageous characteristics will be highlighted.

4 Implementation mechanisms

The presence of generic ‘advantageous characteristics’ also provides a basis for considering the impact of different implementation mechanisms on outcomes. In the case of total purchasing, the pilots laboured in isolation to use the new structural arrangements open to them. In the absence of central support, each pilot had to invent its own wheels. In contrast, NPAT provided considerable resources to the booking pilots to support local management and share relevant ideas and skills. What impact did this additional mechanism have? Did pilots with advantageous characteristics gain more from NPAT’s support than other pilots (by being able to more rapidly implement apposite advice), or was the opposite the case (pilots with advantageous characteristics knew what they were doing already, and NPAT’s support enabled the other pilots to ‘catch up’)? Similarly, the more recent use of the breakthrough collaborative model (Bate et al, 2002; Robert et al, forthcoming) differs from the booking approach by, in particular, its basis on Nolan’s ‘model for improvement’ (Langley et al, 1996)[31].

The evaluations of the NBAP and the CSC have tended to view the impact of NPAT/MA as a constant ‘black box’ not affecting the range of the pilots’ performance, which represents a weakness in terms of Pawson and Tilley’s approach.

5 Outcome measurement

Despite the intentions of the Modernising government White Paper (Prime Minister and the Minister for the Cabinet Office, 1999), policy pilot programmes in the NHS are not required to facilitate outcome assessment by, for example, reporting a minimum level of analysis. Outcomes, if reported at all, usually take the form of selected ‘highlights’ from a minority of participants. Any locally collected and analysed data tend not to be comparable, with a process of standardisation being seen as a benefit for subsequent waves.

In this context, data collection for the purpose of an evaluation is necessarily an additional burden on pilots. The current trend for data collection to depend on the goodwill of a busy project manager provides an inadequate basis for outcome assessment. Evaluation commissioners should move more rapidly than is customary to address the issue of data collection. In the absence of relevant routine data, evaluation commissioners should be prepared if necessary to fund data collection in support of adequate outcome assessment. In general, greater effort should be attached to agreeing objectives, and the consequences for outcome assessment, by commissioners, pilots and evaluators.

Recent experience shows that a lack of basic data can result in the insight into the impact of major policy initiatives being highly constrained (Bate et al, 2002; Robert et al, forthcoming).

---

[31] As a direct offshoot of the NBAP, the phase I cancer services collaborative (CSC) provided, in theory, an ideal opportunity to consider this innovation. In practice, outcome data available to the evaluation of the CSC were so limited that it was not possible to satisfactorily link pilot-level outcomes to advantageous characteristics (Robert et al, forthcoming).

References

Audit Commission (1996) *What the doctor ordered.* HSMO, London

Barnes P, Emerson P, Hajnal S, Radford W, Congleton J. (2000) Influence of an anaesthetist on nurse-led, computer-based, pre-operative assessment *Anaesthesia* 55; 576-589

Bate S, Robert G, McLeod H. (2002) *Report on the 'Breakthrough' Collaborative approach to quality and service improvement within four regions of the NHS* Birmingham: HSMC

Bensley D, Halsall J, McIlwain C, Scott L. (1997) *Total booking systems for elective admission: feedback for site visits* Department of Health internal report

Berwick D (1996) A primer on leading the improvement of systems. *British Medical Journal* 312:619-22

Berwick D, Godfrey A, Roessner J. (1990) *Curing Health Care: new strategies for quality improvement.* San Francisco, CA, Jossey-Bass

Black N (2001) Evidence based policy: proceed with care. *British Medical Journal* 323:275-9

Bosanquet N, Jarman B, Dolan S, Watson L, Bruster S. (1996) *The Bromsgrove Total Purchasing Project 1994-1996.* London: Imperial College School of Medicine at St Mary's

Bowns I, McNulty T. (1999) *Re-engineering Leicester Royal Infirmary: an independent evaluation of implementation and impact.* Sheffield, School of Health and Related Research, University of Sheffield

Department of Health (1998a) *GPs to book patients' hospital appointments on the spot* Press Release, 98/297

Department of Health (1998b) *Pilots announced for booking on the spot hospital appointments* Press Release, 98/401

Department of Health. (2002a) *Primary Care Act Personal Medical Services (PMS) Pilots*. http://www.doh.gov.uk/pricare/pca.htm (accessed December 2002)

Department of Health. (2002b) *NHS Dentistry: Options for Change*. London, Department of Health

Department of Health. (2002c) *Policy Research Programme, Changing Workforce Programme, Research Specification*. London, Department of Health

Enthoven A. (1985) *Reflections on the management of the National Health Service.* London, The Nuffield Provincial Hospitals Trust

Enthoven A. (1999) *In Pursuit of an Improving National Health Service.* London, The Nuffield Trust

Fearn E, Bensley D. (2000) *National Booked Admissions Baseline Survey - Draft Report*. Department of Health internal report

Ham C, Hunter D, Robinson R (1995) evidence based policy making British Medical Journal

Ham C, Kipping R, McLeod H, Meredith P. (2002) *Capacity, Culture and Leadership: lessons from experiences of improving access to hospital services. Final report from the evaluation of the National Booked Admissions Programme first wave pilots* Birmingham: HSMC

Harrison S, Wood B. (1999) Designing health service organization in the UK, 1968 to 1998: from blueprint to bright idea and 'manipulated emergence'. *Public Administration* 77:4;751-768

Hill K, Goodwin N, Morris J, Hall A, McLeod H, Burke T. (forthcoming) *Final report of the national evaluation of Personal Dental Services (PDS)* Birmingham, University of Birmingham

Houghton P, Brodribb A. (1989) Failure to attend for operation: a comparison between admissions and the waiting list system *British Medical Journal* 299; 1139-40.

Joss R, Kogan M. (1995) *Advancing quality: total quality management in the National Health Service* Buckingham, Open University Press

Judge K. (2000) Testing evaluation to the limits: the case of English Health Action Zones *J Health Serv Res Policy* 5:1;3-5

Kasper J, Plume S, O'Connor G. (1992) A Methodology for QI in the Coronary Artery Bypass Grafting Procedure Involving Comparative Process Analysis. *Quality Review Bulletin* April; 129-33

Kilo C. (1998) A framework for collaborative improvement: lessons from the Institute for Healthcare Improvement's breakthrough series. *Quality Management in Health Care* 6:4;1-13

Kilo C. (1999) Improving care through collaboration. *Pediatrics* 103:1;384-393

Kipping R, Meredith P, McLeod H, Ham C. (2000) *Booking Patients for Hospital Care: a progress report. Second interim report from the evaluation of the National Booked Admissions Programme First Wave Pilots.* Birmingham, HSMC

Langley G, Nolan K, Nolan T, Norman C, Provost L. (1996) *The improvement guide: a practical approach to enhancing organizational performance* San Francisco, CA, Jossey-Bass

Layton A, Moss F, Morgan G. (1998) Mapping out the patient's journey: experiences for developing pathways of care. *Quality in Health Care* 7(Suppl):S30-S36

Malone G. (1995) Dental Services *Hansard* 20 November: Column: 28 http://www.parliament.the-stationery-office.co.uk/pa/cm199596/cmhansrd/vo951120/text/51120w05.htm (accessed October 2002)

Mays N, Wyke S, Evans D. (2001a) The evaluation of complex health policy: lessons from the UK Total Purchasing Experiment. *Evaluation* 7:4;405-426

Mays N, Wyke S, Malbon G, Goodwin N. editors. (2001b) *The purchasing of health care by primary care organisations. An evaluation and guide to future policy*. Buckingham: Open University Press

McLeod H, Raftery J. (2000) *Total purchasing and the management of emergency hospital activity.* Total Purchasing National Evaluation Team Working Paper. London, King's Fund

NHS Executive (1994) *Developing NHS purchasing and GP fundholding: towards a primary care-led NHS* EL(94)79 Leeds, NHS Executive

NHS Executive. (1999) *National Booked Admissions Programme: 1999/2000* HSC 1999/102 London: Department of Health,.

NHS Executive. (2000) *National Booked Admissions Programme: 2000/01* HSC 2000/15 London: Department of Health

NHS Executive. (2001) *Moving to mainstream* HSC 2001/009 London: Department of Health

NHS Management Inquiry (1983) *Report*. London, DHSS

NHS Modernisation Agency (2002a) *Field site application form and guidance Modernising Dentistry Programme.* London, NHS Modernisation Agency

NHS Modernisation Agency (2002b) *Improving cancer services: progress report from the cancer services collaborative.* London, NHS Modernisation Agency

National Health Service (Primary Care) Act 1997 London, The Stationary Office

O'Connor G, Plume S, Olmstead E, Morton J, Maloney C, Nugent W et al. (1996) A regional intervention to improve the hospital mortality associated with coronary artery bypass graft surgery. *JAMA* 275:841-6

Packwood T, Pollitt C, Roberts S. (1998) Good medicine? A case study of business process re-engineering in a hospital. *Policy and Politics* 26:4;401-415

Pawson R, Tilley N. (1997) *Realist evaluation.* London, Sage Publications

Pettigrew A, Ferlie E, McKee L. (1992) *Shaping strategic change. Making change in large organisations: the case of the National Health Service.* London, Sage Publications

Plsek P. (1997) Collaborating across organizational boundaries to improve the quality of care. *American Journal of Infection Control* 25:85-95

Plsek P, Wilson T. (2001) Complexity, leadership, and management in healthcare organisations. *British Medical Journal* 323:746-9

PMS national evaluation team (2002) *National evaluation of first wave NHS personal medical services pilots.* Manchester, NPCRDC, University of Manchester

Prime Minister and the Minister for the Cabinet Office (1999) *Modernising government.* Cm 4310 London, The stationery Office

Regen E, Smith J, Goodwin N, McLeod H, Shapiro J. (2001) *Passing on the baton: final report of a national evaluation of primary care groups and trusts.* Birmingham, HSMC

Regen E, Smith J, Shapiro J. (1999) *Lessons from GP commissioning pilots for primary care groups.* Birmingham, HSMC

Robert G, McLeod H, Ham C. (forthcoming) *Evaluating phase I of the NHS Cancer Services Collaborative: lessons for modernising cancer services* Birmingham, HSMC

Royal College of Surgeons of England. (1991) *Guidelines for the Management of Surgical Waiting Lists.* London, The Royal College of Surgeons of England

Secretary of State for Health (1996) *Primary care: delivering the future* Cm 3512. London, The Stationery Office

Secretary of State for Health. (2000) *The NHS Plan A plan for investment A plan for reform.* London: Stationery Office

Secretaries of State for Health, Wales and Scotland (1989) *Working for Patients* Cm 555. London, HSMO

Simpson D, McCallum D. (1992) Involving the Patient Can Free Resources *Health Bulletin* 50/4 July; 288-91.

Smith J, McLeod H, Goodwin N, Morris J, Hill K, Hall A, Bradnock G, Ham C. (2000) *New models of NHS dental care: interim report of the national evaluation of Personal Dental Services* Birmingham, HSMC

Southam J, Talbot R. (1980) Planned surgical admissions in a district hospital *British Medical Journal* 808-9.

South Western Regional Health Authority (1991) *Providing high quality services: a booking approach for elective surgical admissions.* South Western Regional Health Authority,.

Walsh N, Andre C, Barnes M, Huntington J, Rogers H, Hendron C, McLeod H. (2002) *First wave PMS Pilots: Opening Pandora's Box* Birmingham: HSMC

Walsh N, Shapiro J, Davidge M. (1999) *The Berkshire Integrated Purchasing Project – the long and winding road.* Birmingham: Health Services Management Centre, University of Birmingham

Wilkin D, Gillam S, Coleman A. (eds) (2001) *The national tracker survey of primary care groups and trusts 2000/2001: modernising the NHS?* Manchester, NPCRDC, University of Manchester

Wistow G, Harrison S. (1998) Rationality and rhetoric: the contribution to social care policy making of Sir Roy Griffiths 1986-1991. *Public Administration* 76:Winter;649-668

Yates J (1998) *Do booking systems and scoring systems have a contribution in the UK?* Birmingham University: Health Services Management Centre