

Assessing Benefits from Department of Health and National Health Service Research & Development

Martin Buxton, Steve Hanney, Tim Packwood, Simon Roberts and Penny Youll

Public services in the UK are increasingly expected to account for their outputs and performance. This article describes a retrospective evaluation of the benefits from the R&D funded by a regional office of the NHS Executive. The methods adopted enabled the various elements of the region's R&D portfolio to be examined and provided a basis for the development of a plan for regular monitoring. The proposals for implementation are now also feeding into a debate as to how health services R&D might best be monitored nationally in a way that is compatible with the norms and practices of research in other contexts.

Public services in the UK are under increasing pressure to account for their outputs and performance (Treasury Select Committee, 1999; McConaghy, 1999). The first Public Service Agreements (PSAs) produced by UK government departments (Chief Secretary to the Treasury, 1998) partly exempted research funding bodies from the quantitative performance targets applied to some public service delivery organizations. However, they are vulnerable to growing demands that they demonstrate their utility, which might not be as self-evident as is the case for many public services. A sensible approach, therefore, is for public research funding bodies to develop their own system for assessing the benefits from research. They have long used peer review to evaluate the scientific quality of research proposals and reports, but now need to demonstrate the benefits, or payback, from the research they fund.

Such tensions are not unique to the UK. The US National Institutes of Health (NIH) produced a series of examples of the value of its research in terms of how it has resulted in cost savings—both in the form of reduced treatment costs and through enabling treated patients to return to work (NIH, 1993). However, American research funding bodies have successfully argued the case to be allowed to use the alternative format of descriptive performance goals rather than the quantitative goals that apply to many other agencies under the 1993 Government Performance and Results Act (Cozzens, 1997; NSF, 1998).

This article describes an assessment of the benefits from the R&D Programme of the NHS region in the UK, North Thames, which had the largest R&D budget. The aim of the evaluation project was twofold: to provide a substantive

retrospective assessment of the benefits from the North Thames R&D portfolio; and to build on this experience and develop an implementation plan to show how such evaluation, or monitoring, could be undertaken on a regular basis.

Initial decisions were taken about the purposes of the assessment, which categories of benefit should be used, the most appropriate methods to adopt, and how the various elements of the R&D portfolio should be organized for assessment purposes. This article describes the methods used to assess the payback from each element of the R&D Programme.

Planning the Assessment

The main purposes of undertaking a retrospective assessment of the benefits from an R&D programme were:

- To provide evidence that might strengthen the justification for R&D expenditure.
- To assist in prioritizing future investments in R&D.
- To indicate ways of improving the conduct of research, and R&D management, to increase the magnitude or likelihood of subsequent beneficial consequences.

The need for justifying R&D expenditure involves several, possibly contradictory, pressures. In addition to the general pressure to account for all public expenditure, recent changes relating to NHS expenditure on R&D mean that it is now more clearly seen as being in competition with spending directly on health services. On the other hand, the complexities of the processes through which research findings might eventually create impacts on policy and practice mean that many have argued that such outcomes are limited

Martin Buxton and Steve Hanney are in the Health Economics Research Group, Brunel University.

Tim Packwood is in the Department of Government, Brunel University.

Simon Roberts is at the Centre for Research in Social Policy, Loughborough University.

Penny Youll is in the Evaluation and Development Research Unit, the Tavistock Institute.

and difficult to trace (see, for example, Weiss, 1980). As a consequence some have questioned the appropriateness of attempting to make impact assessments. However, there is now 'an unusual opportunity for research evidence to have substantial impact on policy and practice' (Davies *et al.*, 1999, p. 3), and, furthermore, 'Evidence on effectiveness is more to the fore in health care than in any other public service' (Davies and Nutley, 1999, p. 15). There is, therefore, now perhaps both a greater need and opportunity to assess the payback from health research.

Next, the methods were clarified. We started with the existing multidimensional categorization of benefits devised for the Department of Health by the Health Economics Research Group (HERG) (Buxton and Hanney, 1994, 1996 and 1997). The categories include both outputs and outcomes and figure 1 represents an updated version of the original categorization. A sub-category for organizational development has been added to reflect the fact that community-based agencies and voluntary organizations, as a result of their participation in research, may experience increased funding and resources, an extension of networks, higher local or national status, or other organizational development. This may then lead to an eventual enhancement in the services provided.

The multidimensional approach can be criticised on the grounds that it may lead to double counting of benefits. However, an earlier review (Buxton and Hanney, 1998) of single all-embracing measures suggested that each approach had fundamental weaknesses which outweighed the apparent analytical advantage of a single metric. These disadvantages include the limited coverage of the payback categories involved in bibliometric and cost savings approaches (see, for example, Anderson, 1989; Detsky, 1989; NIH, 1993) and the immense difficulty of putting a monetary value on all benefits.

Buxton and Hanney (1996, 1997) used a sequential model to characterize the process whereby R&D generates outputs and final outcomes. It starts with the identification of R&D needs. Then there is the first of two interfaces between the political and professional environment and the research processes. The permeability of the interfaces (Kogan and Henkel, 1983) can be important if outcomes are to be achieved and the first interface involves the commissioning and specification of research to meet the identified need. The primary, or initial, outputs emerge from the subsequent research processes and include publications and trained researchers.

There is then the second, or dissemination, interface when findings are relayed to the political or professional environment. This is followed by a stage which may include the generation of secondary outputs in the form of policy statements or administrative decisions (manifest in, for

example, national ministerial decisions or local guidelines) and by an applications stage when research findings are put into practice. Final outcomes result from this and consist of both health service related and more general economic benefits. Although in its simplest form this is a linear model, it can incorporate and highlight the variety of possible feedbacks that occur, and can be contextualized within a model recognizing the existence of a 'knowledge pool' (Gibbons and Johnston, 1974) into which research findings are fed and in which ideas and people interact (Cozzens, 1997; Hanney *et al.*, 2000). The existence of such a knowledge pool, and the fact that impacts of research are often unpredictable and spread over a long period (see, for example, Weiss, 1980), highlight the fact that it is desirable for any assessment of R&D to recognize the norms of research and not attempt to make use of precisely quantifiable targets.

Assessing the outcomes from the provision of services such as health is itself a most difficult task (see, for example, Smith, 1995). But when assessing the outcomes of research it is necessary to identify the impact the research has had on policy and practice even before attempting to assess the benefits in terms of improvements in the service provided.

Experience has shown case studies to be very illuminating when conducting payback assessments (Buxton and Hanney, 1994; Williams and Rank, 1998; Bozeman and Klein, 1999). Detailed case studies are, however, too resource-intensive to be used for comprehensive regular monitoring, so a series of questionnaires was devised for the bulk of the assessment of North Thames R&D with case studies used sparingly.

The evaluation focused on the four elements of the North Thames R&D portfolio. The field work was conducted in late 1997 and the first half of 1998. An account of the methods, and an outline of the findings, for each element is contained in the following four sections of the article but a fuller account is available in the report to North Thames (Buxton *et al.*, 1999).

The Projects

The lead applicants for all projects funded in the 1990s, which had been completed by 1997, were sent a questionnaire to see how far this could provide similar information about benefits to that gathered by the team in previous studies using more intensive case study techniques. (Details of many of the North Thames projects, which range from health services research to more basic biomedical research, may be found on the R&D Directorate's web site: www.doh.gov.uk/ntrd/rd.htm). The design of the questionnaire was informed by the HERG payback categorization and model. Questionnaires were issued for 164 projects and 115 (70%) were returned.

The average number of publications of any sort per project was 2.2: of these just under half were journal articles and they came from 55

Figure 1. Categories of payback.

(a) *Knowledge*

(b) *Benefits to future research and research use:*

- (i) The better targeting of future research.
- (ii) The development of research skills, personnel and overall research capacity.
- (iii) A critical capability to utilise appropriately existing research including that from overseas.
- (iv) Staff development/educational benefits.

(c) *Political and administrative benefits:*

- (i) Improved information bases on which to take political and executive decisions.
- (ii) Other political benefits from undertaking research.

(d) *Health sector benefits:*

- (i) Cost reduction in the delivery of existing services.
- (ii) Qualitative improvements in the process of service delivery.
- (iii) Increased effectiveness of services.
- (iv) Equity, for example improved allocation of resources at an area level, better targeting and accessibility.
- (v) Revenues gained from intellectual property rights.
- (vi) Organizational development.

(e) *Broader economic benefits:*

- (i) Wider economic benefits from commercial exploitation of innovations arising from R&D.
- (ii) Economic benefits from a healthy workforce and reduction in working days lost.

Source : Adapted from Buxton and Hanney (1996).

projects. Citations analysis has had only a limited role in the formal assessment of research (Cave *et al.*, 1995; Cave *et al.*, 1997); nevertheless, it was considered useful to examine how it changed the picture given by a simple count of publications. An analysis was undertaken using the relevant journal impact factor scores as measured by the Institute for Scientific Information (ISI, 1997). Our analysis confirms, however, that impact factor analysis may underestimate the payback from some applied projects in that no account is taken of some journals which nevertheless may be widely read.

Over a quarter of projects contributed in some way to a post-graduate qualification, including 10 MDs and 12 PhDs. For the post-graduate qualifications, as with the publications, there are, however, problems in assessing the precise contribution of the specific North Thames funding. In some instances, the qualifications represent outputs from a stream of research funded sequentially or simultaneously from several sources.

Approximately 40% of respondents indicated that their research had already made an impact in terms of one, or more, of:

- Policy.
- Practice.
- Health benefits.

Over 60% expected it to do so. Overall, 80% of respondents replied Yes to having had an actual or expected benefit under at least one of these three headings. A scoring system was devised and tested to give an indicator of magnitude of

benefit represented by the more detailed written responses about impact on policy, practice and health benefit. The mean of the two assessors' scores was used as an indicator of 'benefit' in subsequent analyses undertaken to see whether there were any strong relationships between key variables. A major finding is that the bibliometric measures were not well correlated with our 'benefit' score. Indeed almost half the projects that did not have any journal publications had a benefit score, indicating that they appeared to have made some impact on policy or practice through dissemination by means other than journal publication and/or through being applied in the unit in which the research was conducted. (See Hanney *et al.*, 1999, for a more detailed discussion of the scoring system, correlations between variables and the case studies discussed below.)

A sample of the projects was selected for further assessment as case studies. One case study revealed that the questionnaire had overstated the level of local application of the findings from the project, but understated the impact they might be having on guidelines at an international level. Overall, the comparison of questionnaire based benefit score and case study based re-score showed that on average, for this subset of projects, the questionnaires slightly understated the impact. This does not lend support to the presumption, often made, that questionnaire responses would exaggerate the impact of research. Therefore, from comparative analysis of the data from the questionnaires and from the case studies, we conclude that a well

designed questionnaire might be a useful and sufficiently reliable technique to obtain a broad-brush picture of a set of projects, even though case studies provide a fuller and more reliable basis for direct comparison of individual projects.

The Mental Health Programme

An assessment was made as to whether a commissioned and themed R&D programme of research provides added benefits beyond those of individual projects. A higher proportion of projects from the chosen programme, Mental Health (MHP), were subject to case studies and assessments were made of the programme as a whole. The five MHP case studies confirmed indications of emerging payback from the projects. The case studies highlighted several issues more clearly than the questionnaires had done, including the importance of community-based services and agencies in formative work between researchers and providers.

To obtain an indication of knowledge about, and utilization of, the findings of the MHP projects a short questionnaire was sent to the person with lead responsibility for mental health in each health authority and trust in the region. From a response rate of over 60%, the majority indicated that they had no knowledge of the projects or their findings. This might suggest that, for example, consideration could be given to the possible desirability of funding a programme director for each R&D programme, as does the Economic and Social Research Council, who would have responsibility for dissemination and networking with policy-makers and practitioners.

The overall assessment is that, while the individual projects were selected with considerable care and are delivering benefits broadly in line with their objectives, there was no clear evidence that the programme as a whole had provided added value to the processes, outputs, or outcomes, of the individual projects taken separately.

Centres

The category of 'centres' covered ten units funded in various ways and for diverse purposes. The two centres evaluated varied in size, history and scope. The evaluation was not conceived as an attempt directly to compare the two. Nevertheless, the results of the analysis were in line with these differences in background. For each centre we examined their activity as a whole, and selected several individual projects for more detailed examination through the usual case study techniques of documentary analysis and interviews. The case studies were structured according to the evaluative framework described earlier and an analysis of how far such methods are compatible with broader developments in the evaluation of health research centres is reported in Hanney *et al.* (2000).

For the larger Centre One, substantial evidence was gathered that indicated payback

had been achieved across the range of categories, and in particular in generating knowledge and influencing policy, with one respondent suggesting it was 'a place where policy-makers would go'. This, in turn, had impacted on health. The smaller, and more recently established, Centre Two was part of the national programme to strengthen R&D in Community and Primary Care. Inevitably its achievements were more limited and concentrated in the early stages of the HERG model and were concerned with determining research needs and developing research capabilities.

Not surprisingly, Centre One had a far stronger centre identity than Centre Two and various benefits that can flow from funding a centre were demonstrated. These include the way a centre can draw on its own 'reservoir of knowledge' to help set a research agenda and better target future research, and the way it can develop a network of links that enhance the likelihood of the research impacting on policy.

The Education and Training Programme

The North Thames schemes for PhD fellowships and masters studentships are designed to increase research capacity. This is beginning to occur. Working with the E&T co-ordinator appointed by North Thames, it was possible to elaborate the existing questionnaires used to survey students. As a result, it can be shown that some of the research undertaken by the fellows is already impacting on policy and practice and that some masters students are not only becoming involved in research themselves but are also encouraging their professional colleagues to be positive towards evidence-based health care (EBHC).

The North Thames programme of EBHC courses was designed to promote research awareness and the utilization of research findings. It was intended to reach a wide compass of personnel in the NHS—community as well as hospital based managers and clinicians, and all disciplines. The evaluation of the EBHC courses included a detailed analysis of the developmental processes which resulted in a framework for evaluating the payback from educational opportunities being devised. In figure 2, A–D represent the sequence of phases involved in the translation of knowledge and skills into health gains. Often, educational courses are evaluated only in terms of transmission or delivery (A), i.e. the way a course was run, its content, the quality of teaching etc. But this does not address issues of whether individuals actually acquire and use new ideas (etc.) in the workplace. The framework suggests that before actual gains (D) can be recognized and attributed to any specific educational opportunity, it must first be demonstrated that knowledge and skills have been transmitted (A); that these have been acquired (B); and put into practice (C) by the individual. The measurement of health gain from

Figure 2. Framework for evaluating educational opportunities in the NHS.

	A	B	C	D
<i>Evaluative focus:</i>	Transmission/delivery	Acquisition of knowledge/skills	Application of knowledge/skills through policy-making and in practice	Health gains
<i>Context or setting:</i>	Educational opportunity (course, workshop etc.)	The individual as learner	Workplace	Patient/patient care
<i>Measures:</i>	Relevance/quality of content and teaching. Support for application of learning	Gains in knowledge/skills. Commitment to application of learning	Evidence of application/changes in practice	Changes in patient health and/or service delivery

investment in education and training will be dependent on how clearly connections can be made between the course and changes in practice and/or policy.

Using this framework, the analysis included a survey of all participants in a one-week workshop 17 months after their attendance. Nearly all the 50% who responded were positive about the workshop and EBHC, and indicated it had promoted their utilization of EBHC in the workplace. In addition, an examination of a collaborative approach between a course provider and trusts indicated that important workplace development may be stimulated given the right preparation and support. There are indications that specific projects initiated through EBHC course participation might provide quite concrete evidence of applications and of changes in clinical practice.

Conclusions from the Assessment

Turning to the three purposes of assessing retrospectively the benefits from R&D, evidence was provided that could be used to strengthen the justification for expenditure on R&D. Aggregate analyses provided some quantitative indicators and individual case studies provided in-depth examples that illustrated the nature, and complexity of the processes. Furthermore, although the analysis suggests the impact of research varies greatly with a range of factors, including the nature of the research and findings and the political salience of the subject, examples were found, from centres and from individual projects, where there was evidence that impact had been rapid.

While the evaluation did not provide a clear guide for prioritizing future R&D investment, it did indicate ways in which the conduct of research

and research management might be improved, and confirmed earlier findings that if payback is to be maximized then time and effort has to be invested.

Discussion

This evaluation underlined the importance of the clarification of the objectives of each element of an R&D Programme and the desirability of a coherent evaluative strategy linked to these objectives. A series of specific recommendations was made to North Thames concerning the methods that might be applied routinely in assessing benefits from the regionally funded activities. It proposed appropriate methods of assessment for each of the main elements of the R&D programme, and provided draft questionnaires where necessary for application to projects and to individual researchers in receipt of training awards (for details see Buxton *et al.*, 1999).

There is now a move at national level towards developing a regular monitoring system for all Department of Health/NHS R&D, including the money that goes directly to hospital trusts (Buxton, Croxson and Hanney, 1999). The North Thames assessment offers a number of lessons. In terms of a multidimensional categorization, the evidence from the projects confirmed that publications on their own do not provide a sufficient indicator of broader impacts on policy or health benefits. As regards the methods to use in payback analysis, the North Thames study indicates the potential feasibility of gathering information directly from researchers about a range of payback categories. Case studies, as previously used by HERG, were shown still to have an important role in providing good stories to justify R&D expenditure. Furthermore they

provide some verification of the information in self-completed questionnaires and, by highlighting certain issues, they can indicate ways in which the questionnaires could be improved. External data sets such as citation analyses were found to provide another important source of information, but the study also showed that there are limitations on their application. User surveys can also provide valuable data. Looking more widely, the usefulness of the HERG approach has also been discussed in relation to educational R&D (Kogan, 1997).

Within the context of a general move towards greater performance measurement in the public services it is important for research funding bodies to develop techniques to show they are responsive to the political imperative. It is also desirable for them, however, to retain the flexibility to undertake performance assessment in a way that is compatible with the norms and practices of research. The proactive development of specific systems for monitoring R&D payback is probably the best way to achieve these aims. ■

Acknowledgements

This work was funded by the North Thames NHS Executive and by the Department of Health's Policy Research Programme. The authors thank those many researchers in North Thames who completed questionnaires and co-operated with case studies. Responsibility for the text remains solely that of the authors.

References

- Anderson, J. (1989), *New Approaches to Evaluation in UK Research Funding* (Science and Policy Support Group, London).
- Bozeman, B. and Klein, H. (1999), The case study as research heuristic: lessons from the R&D value mapping project. *Evaluation and Program Planning*, 22, pp. 91–103.
- Buxton, M. and Hanney, S. (1994), *Assessing Payback from Department of Health Research and Development. Vol. 1: The Main Report*. Research Report No. 19 (HERG, Brunel University, Uxbridge).
- Buxton, M. and Hanney, S. (1996), How can payback from health services research be assessed? *Journal of Health Services Research and Policy*, 1, pp. 35–43.
- Buxton, M. and Hanney, S. (1997), *Assessing Payback from Department of Health Research and Development: Second Report. Vol. 1: Main Report*. Research Report No. 24 (HERG, Brunel University, Uxbridge).
- Buxton, M. and Hanney, S. (1998), Evaluating the NHS R&D programme: will the programme give value for money? *Journal of the Royal Society of Medicine*, 91 (Suppl. 35), pp. 2–6.
- Buxton, M., Croxson, B. and Hanney, S. (1999), *Assessing the Payback from Health R&D: From 'ad hoc' Studies to Regular Monitoring*. Research Report No. 27 (HERG, Brunel University, Uxbridge).
- Buxton, M., Hanney, S., Packwood, T., Roberts, S. and Youll, P. (1999), *Assessing the Benefits from North Thames Research and Development*, Research Report No. 25 (HERG, Brunel University, Uxbridge).
- Cave, M., Hanney, S. and Henkel, M. (1995), Performance measurement in higher education—revisited. *Public Money & Management*, 15, 4, pp. 17–23.
- Cave, M., Hanney, S., Henkel, M. and Kogan, M. (1997), *The Use of Performance Indicators in Higher Education* (3rd edn) (Jessica Kingsley, London).
- Chief Secretary to the Treasury (1998), *Public Services for the Future: Modernization, Reform, Accountability—Comprehensive Spending Review: Public Service Agreements 1999–2000*. Cm 4181 (The Stationery Office, London).
- Cozzens, S. E. (1997), The knowledge pool: measurement challenges in evaluating fundamental research programs. *Evaluation and Program Planning*, 20, pp. 77–89.
- Davies, H. T. O., Nutley, S. M. and Smith, P. C. (1999), Editorial: What works? The role of evidence in public sector policy and practice. *Public Money & Management*, 19, 1, pp. 3–5.
- Davies, H. T. O. and Nutley, S. M. (1999), The rise and rise of evidence in health care. *Public Money & Management*, 19, pp. 9–16.
- Detsky, A. (1989), Are clinical trials a cost-effective investment? *Journal of the American Medical Association*, 262, pp. 1795–1800.
- Gibbons, M. and Johnston, R. (1974), The roles of science in technological innovation. *Research Policy*, 3, pp. 220–242.
- Hanney, S., Davies, A. and Buxton, M. (1999), Assessing benefits from health research projects. *Research Evaluation*, 8, 3, pp. 189–199.
- Hanney, S., Packwood, T. and Buxton, M. (2000), Evaluating the benefits from health R&D centres: a categorization, a model and examples of application. *Evaluation*, 6, 2, pp. 137–160.
- Institute for Scientific Information (1997), *Journal Citations Reports: A Bibliometric Analysis of Science Journals in the ISI Database* (ISI, Philadelphia, PA).
- Kogan, M. (1997), Learning from other areas of study. In Hegarty, S. (Ed), *The Role of Research in Mature Education Systems: Proceedings of the NFER International Jubilee Conference* (NFER, Slough).
- Kogan, M. and Henkel, M. (1983), *Government and Research: The Rothschild Experiment in a Government Department* (Heinemann, London).
- McConaghy, D. (1999), Measuring (accountable) success—analysis of the government's measures of success. *Public Money & Management*, 19, 4, pp. 62–65.
- National Institutes of Health (1993), *Cost Savings Resulting from NIH Research Support* (NIH, Bethesda, MD).
- NSF (1998), *National Science Foundation: FY 1999 GPRA Performance Plan*. (See <http://www.nsf.gov/pubs/1998/nsf99gprapp/start.htm>).
- Smith, R. (1995), Performance indicators and outcome in the public sector. *Public Money & Management*, 15, 4, pp. 13–16.
- Treasury Select Committee (1999), *Public Service Agreements, Seventh Report of Session 1998–99*. HC378 (The Stationery Office, London).
- Weiss, C. (1980), Knowledge creep and decision accretion. *Knowledge: Creation, Diffusion, Utilization*, 1, 3, pp. 381–404.
- Williams, A. and Rank, A. R. (1998), Measuring the economic benefits of R&D. *Research Evaluation*, 7, 1, pp. 17–30.