Assessing the impact of health technology assessment in the Netherlands

Wija J. Oortwijn
ECORYS Nederland BV

Stephen R. Hanney
Brunel University

Andreas Ligtvoet
Delft University of Technology

Stijn Hoorens, Steven Wooding, Jonathan Grant
RAND Europe Cambridge Ltd.

Martin J. Buxton
Brunel University

Lex M. Bouter
VU University Amsterdam

Objectives: Investments in health research should lead to improvements in health and health care. This is also the remit of the main HTA program in the Netherlands. The aims of this study were to assess whether the results of this program have led to such improvements and to analyze how best to assess the impact from health research.
Methods: We assessed the impact of individual HTA projects by adapting the "payback framework" developed in the United Kingdom. We conducted dossier reviews and sent a survey to principal investigators of forty-three projects awarded between 2000 and 2003. We then provided an overview of documented output and outcome that was assessed by ten HTA experts using a scoring method. Finally, we conducted five case studies using information from additional dossier review and semistructured key informant interviews.
Results: The findings confirm that the payback framework is a useful approach to assess the impact of HTA projects. We identified over 101 peer reviewed papers, more than twenty-five PhDs, citations of research in guidelines (six projects), and implementation of new treatment strategies (eleven projects). The case studies provided greater depth and understanding about the levels of impact that arise and why and how they have been achieved.
Conclusions: It is generally too early to determine whether the HTA program led to actual changes in healthcare policy and practice. However, the results can be used as a baseline measurement for future evaluation and can help funding organizations or HTA agencies consider how to assess impact, possibly routinely. This, in turn, could help inform research strategies and justify expenditure for health research.

Key words: Impact, Health technology assessment, The Netherlands

Assessing the returns from research is of increasing interest to research-funding organizations (1;6;10;11;13–16). From an international perspective health services are increasingly emphasizing the need to ensure that health care is evidence-based. For those involved in healthcare-related research, including health technology assessment (HTA), the focus lies on providing a strong and robust research foundation and facilitating the translation of research into practice, so ensuring that investments in health research lead to improvements in health and health care. Such improvements are the remit of the main HTA program in the Netherlands.

This Health Care Efficiency Research program is coordinated by the Netherlands organization for health research and development (ZonMw). The program actively promotes research on the recognition, assessment, and implementation of cost-effective interventions and fosters generalization of knowledge with an annual budget of 12.2 million Euro.

In 2000, the program was implemented and commissioned by the Dutch Minister of Health, Welfare, and Sport. In 2001, the first grants for this program were allocated. There are two main ways to support research within this HTA program:

- **Bottom-Up**: In 2001, the Ministry did not specify priorities and an open submission of research proposals resulted. The second round (2002) was a so-called "steered bottom-up approach," which means that, in the event of equal scientific quality, preference was given to priority areas defined by the Ministry of Health, Welfare, and Sport. The priority areas were detection of cancer metastases, mental and behavioral disorders, care of chronically ill, clinical genetics, and infectious diseases.

- **Top-Down**: The focus was on two predefined subjects: positron emission tomography (PET) diagnostics and treatment of fertility disorders in the rounds 2001 and 2002. Almost all applications come from academic hospitals, and many involve randomized clinical trials (RCTs) concerning therapeutic or diagnostic procedures (17;18).

Whether the results of this program have actually led to improvements in health and health care has never been researched in a systematic way. With this study, we aim to contribute to the systematic analysis of the Dutch HTA program by applying a model to assess impact based on the output and outcome of individual HTA projects. We based our approach on a method implemented by HERG, Brunel University, and the Wessex Institute, University of Southampton to assess the payback of HTA projects in the United Kingdom (8). The study described here was also intended to test if and how the "payback framework" might prove useful in reviewing the payback of HTA projects in the Netherlands.

METHODS

Development of Evaluation Framework

We developed an evaluation framework to assess payback of individual HTA projects on the basis of desk research. Members of our research team (Brunel University) have been examining how to conceptualize, describe, and assess the "payback" of funded research since 1993. It is important to acknowledge that this is broader than an assessment of the scientific quality of research per se; the impact of research is the key issue. This has resulted in the payback framework that consists of a conceptual categorization of the benefits arising from health services research and a model for assessing "payback" (2;3;10).

The evaluation framework we used consists of two components: (i) A logic model of the research process. Logic models are widely used in evaluation methodology to understand input-process-output relationships and breakdown research programs into parts. The payback model developed by Buxton and Hanney is an example of a logic model and breaks down the process by which (HTA) research is translated into practice, and provides a way to organize the assessment of this multidimensional categorization of benefits from research (Figure 1).

(ii) The definition of evaluation criteria for output and outcome of funded research. We slightly adapted the original payback criteria of the payback framework and used the following categories: Knowledge Production (e.g., number of publications), Research Benefits (e.g., development of research skills, personnel and capacity), Informing Policy (e.g., development of new treatment guidelines), Changing Health Practice (e.g., use of new treatment strategies), and Broader Impact on Health (e.g., economic benefits from a healthy workforce). It was recognized in advance by the research team that it would be difficult to obtain "hard" data for last two criteria due to the fact that the HTA program had only started in 2000.

Dossier Review

Guided by the framework, two researchers conducted dossier reviews of all forty-three projects granted between 2001 and 2003. Most projects took the form of prospective, clinical trials (two-third of the grants), while others, for example, were modeling studies (18). We read the key documents including the original research grant proposals, referees reports, final reports, and information about the program (annual reports, minutes of committee meetings, and so on). We collected and synthesized the information from the dossiers at ZonMw.

Postal Survey

In parallel, we developed a survey that was based on a (postal) questionnaire that Buxton et al. (4) developed, subsequently refined, and also used for assessing the impact of the UK Department of Health HTA Program (8;9). The relevance of the survey to the particular context was tested by an international HTA researcher, who has experience as a reviewer of the ZonMw grants. We sent the postal survey, including data from the dossier review, to all principal investigators to collect the most recent data on their projects. It might have been a good idea to conduct a separate user survey had the resources for the study been much greater.
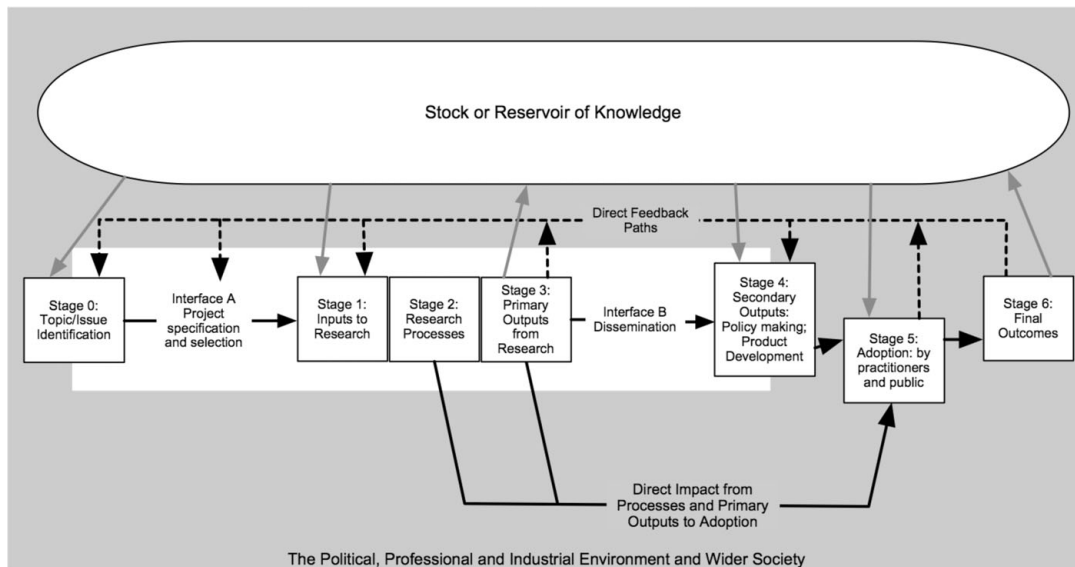
Figure 1. Schematic diagram of the payback model.
The model serves to indicate a series of stages that help to organize payback assessments:
Stage 0: Topic/Issue Identification involves the generation of the original ideas for the research; Interface A: Project Specification and Selection involves the development of research proposals and the (peer) review process;
Stage 1: Inputs to research focus on financial inputs, but also expertise of the research team and knowledge base of the research team;
Stage 2: Research Processes concern how the research was conducted and any difficulties encountered in patient inclusion, appropriateness of methods, involvement of potential users, and so on;
Stage 3: Primary Outputs from Research relate to knowledge production, mainly publications; Interface B: Dissemination reflects the procedure at the end of a study whereby the findings are disseminated to the policymakers and more widely (e.g., by presentations);
Stage 4: Secondary Outputs includes a wide variety of items such as reimbursement decisions by the Ministry of Health; citing of the findings in a clinical guideline from a national or local professional group; or inclusion of the findings in a contract or in a document from an audit, an inspectorate or an evaluative body;
Stage 5: Adoption by Practitioners and Public deals with behavioral change that is necessary to implement the findings of the research into practice to result in final outcomes. Examples are adoption of a new surgical technique that appeared to be cost-effective or introduction of a new pharmaceutical on the market;
Stage 6: Final Outcomes are seen as the ultimate goal of the research. For health research this goal involves improved public health (or quality of life) and broader economic benefits.
Source: (10).

On the basis of the survey data, a first inventory of the outputs and outcomes of the HTA projects was made. The results were presented as a summary; that is, one page standard form about the payback from each HTA study. Using a common structure for all research grants has the advantage of facilitating comparative analysis, allowing us, for example, to identify the factors associated with successful translation of research.

Case Study Analysis

To provide a deeper understanding of how payback occurs, a purposive sample of five HTA projects were investigated in more detail to illustrate useful levels of payback. With the help of the Committee for Health Care Efficiency Research (CDO), we selected for case studies five projects that mirrored the diversity of ZonMw funding. We developed a semistructured interview protocol based upon the protocol used in a previous research impact study (16). In each case study, we explored the initial and longer-term output and outcome of the research in more detail, including the translation of the research (e.g., into practice). We planned to undertake three interviews per case study (either face-to-face or by telephone) including users, the principal investigator, and at least one other involved researcher. In total, we conducted fourteen interviews lasting between 45 and 90 minutes. The interview aimed to verify conclusions from written materials and to identify any payback that might have been missed. In all cases, the interviews were used to update the information retrieved from desk research. Using the information collected from a more in-depth dossier review and semistructured key informant interviews, each of the five cases was written up

as a narrative organized according to the structure provided by the payback model.

Scoring of Payback Using HTA Experts

The scoring of payback is an attempt to standardize the findings from the data gathered with the aim that this will allow comparisons to be made, for example, between different modes of funding HTA projects. To allow such comparisons to be drawn, it was important to include a sufficient number of HTA projects in the scoring exercise. Previous experience of applying the payback framework suggested that at least 18–24 months should elapse between the end of a study and the main payback assessment. This was indeed confirmed by our study, but most of the projects in the Dutch HTA program had not yet reached this time-scale. To generate a sufficient number of projects we, therefore, had to include all the projects that had already led to a final report and for which we had received survey results, even if they had been completed only a short time before the payback assessment was conducted. From the total number of grants ($n = 43$), twenty-six met these two criteria.

The outputs and outcomes of these projects were then reviewed by ten HTA experts, including representatives of patient organizations, health insurers, ministry of Health, health advisory councils, CDO, the Committee for the Evaluation of the Scientific Quality, and health professional bodies. For this purpose, we developed a novel method of scoring the projects in terms of payback that consisted of three rounds of scoring:

First Round Scoring

Information resulting from dossier review and the survey results was used for writing summaries of the impact from the twenty-six HTA projects. We invited sixteen experts to participate in the scoring exercise of which eleven experts indicated their willingness to participate in the study. The experts were asked (by mail) to score the impact of each study based on these summaries and using a scoring system based on those that had been used in previous payback studies (8;16). It provided a method for converting the qualitative descriptions of the payback in each of the payback categories into a quantitative measure. These rating forms were sent—together with scoring instructions—to each expert, after having tested them by two members of the research team. The experts were asked to score each HTA study for all payback categories on a scale ranging from 0 to 10. At the bottom of the scale, 0 represented no payback. At the top end of the scale the value of 10 was a qualitative description of the largest payback in that category that could have been expected (Table 1).

The rating forms of ten experts were returned by post and quantitatively analyzed. It is important to emphasize that the last three scoring categories (Informing Policy, Changing Health Practice, and Broader Impact on Health) differ in a major way from the first two categories (Knowledge Production and Research Benefits) in the sense that the a) and b) scales are related in a different way. In Knowledge Production and Research Benefits, the a) and b) scales cover issues that can be considered separately, and for them, it makes sense just to aggregate the scores. For the other scoring categories, the a) scale in each case covers the scope of the impact (i.e., nature of the policy level, numbers who benefit), whereas the b) scale covers how far the research was responsible. Thus, the combined score for each of these last three categories should be the product of the a) and b) scores (i.e., a × b) not the summation (i.e. a + b) because of their nature ("level" and "degree" of impact) (4). Because of this, we cannot add the ten scores of each project to determine the final payback of each project. We can, however, plot the summarized scores (adding 1a + 1b, 2a + b and multiplying 3a × 3b, 4a × 4b and 5a × 5b) on a spider plot whose size and shape represents the payback from the HTA study.

The scores from the first round were analyzed and sheets showing their distribution were circulated to the experts before an expert meeting (second round scoring). As was done in Wooding et al. (16), we used the RAND/UCLA (University of California at Los Angeles) Appropriateness Method (RAM) for analyzing levels of disagreement and agreement between experts (consistency) (7). The RAM method provides an algorithm for determining whether disagreement exists between scorers. However, this method focuses on the occasions where people are scoring at opposite ends of the scale. Therefore, in addition to the RAM disagreement algorithm we also tested disagreement by checking whether the actual score was more than two or three points from the median score, as was also done in a former impact study (16). To select five HTA projects to be re-scored during the half-day meeting, we used a figure that captures disagreement with regard to the HTA study as a whole. For this purpose, we selected projects with the largest variance of scores, being the sum of the variances of the ten scoring items (criteria).

Second Round Scoring

Nine of eleven experts who were willing to participate in the study were brought together at a meeting to give them the opportunity to discuss the scoring face-to-face, in light of the knowledge of their collective responses and to resolve any misunderstandings about the outputs from the summaries. Each expert was given scoring sheets showing his/her score and the distribution of scores for each criterion and the average score per criterion. At the meeting, we discussed the scoring of the projects which exhibited the largest scoring variance and the arguments used for scoring. We were not necessarily seeking consensus—the expert meeting allowed experts to understand the reasoning that lay behind divergent views. During the expert meeting, we also discussed the methodological approach used. At the end of the discussion, the experts had the opportunity to re-score the projects to determine the areas of final agreement and disagreement.

Table 1. Scoring Scales Payback Categories

KNOWLEDGE PRODUCTION

1a. Presentations
We would like you to consider the number of presentations to international, national or local meetings, where a full presentation is generally ranked higher than poster presentations.

10 – At least 6 presentations made to international conferences.
 8 – At least 3 presentations made to international conferences, or more than 6 presentations made to national meetings.
 6 – At least 1 presentation made to an international conference, or more than 3 presentations made to national meetings.
 4 – More than 1 conference poster or external presentation made, for example to a practitioner, patient, or research group (including continuous medical education).
 2 – At least 1 conference poster or external presentation made, for example to a practitioner, patient, or research group (including continuous medical education).
 0 – No presentations or posters.

1b. Publications
We would like you to consider those publications that contribute to knowledge dissemination; thus, this suggests a judgment of the impact factor and quality of the journals.

10 – The project produced at least 8 peer-reviewed publications (not abstracts or conference proceedings) in respected journals.
 8 – The project produced at least 5 peer-reviewed publications (not abstracts or conference proceedings) in respected journals.
 6 – The project produced at least 3 peer-reviewed publications (not abstracts or conference proceedings) in respected journals.
 4 – The project produced at least 1 peer-reviewed publication (not abstracts or conference proceedings) in a respected journal or a journal relevant for the target audience (not abstracts or conference proceedings), e.g., in a (inter)national professional journal.
 2 – The project produced internal (e.g., report of department) but no external publications.
 0 – The project produced no publications or internal report.

RESEARCH BENEFITS

2a. Capacity building
We would like you to consider the development of research skills, personnel and research capacity, e.g., training of health professionals involved in the research.

10 – The project made a considerable contribution to at least 2 research degrees, such as PhDs.
 8 – The project either made a considerable contribution to at least 1 research degree, or a moderate contribution to at least 2.
 6 – The project made a moderate contribution to at least 1 research degree.
 4 – The project made some contribution to at least 1 research degree.
 2 – The project some contribution to research capacity building in some other way (e.g., training of specific tasks).
 0 – The project made no contribution to research degrees or research capacity building in any other way.

2b. Research targeting
We would like you to consider a variety of contributions to follow-on research: pilot studies, implementation projects, methodological frameworks, or mathematical underpinnings, and so on (absorption of knowledge, including from abroad).

10 – The project made a considerable contribution to more than 1 follow-on project by the team and/or by others.
 8 – The project made a contribution to more than 1 follow-on project, considerable in at least one case.
 6 – The project made a contribution to more than 1 follow-on project, moderate in at least one case.
 4 – The project made a moderate contribution to 1 follow-on project, or any contribution to more than one follow-on project.
 2 – The project made a contribution to at least 1 follow-on project.
 0 – The project made no contribution to targeting of future research.

INFORMING POLICY

We would like you to consider policy making at (inter)national or local level of the health system, e.g., reimbursement decisions by the Ministry of Health, citing of the findings in an advice of the Health Council, citing of the findings in a clinical guideline from a national or local professional group, inclusion of the findings in a contract or in a document from an audit, an inspectorial or an evaluative body).

3a. Nature of the policy (e.g., reimbursement decisions, use in clinical guidelines)
10 – The project made an impact on a substantial policy of an international body or substantial policies of at least the national government.
 8 – The project made an impact on at least 1 policy (brief) from a national body such as the Health Council (GR), Council for Health Research (RGO).
 6 – The project made an impact on at least 1 policy from a national professional body.
 4 – The project made an impact on the policymaking of at least 1 local unit of the health service.
 2 – A claim for impact was made but no details given, or details given of a claim for expected future impacts.
 0 – The project made no impact on policies.

Table 1. Continued.

3b. Degree of impact on health policy

10 – At least 1 of the policies was almost solely based on the project's evidence in a direct instrumental way.
 8 – The project made a considerable impact on the policy/policies.
 6 – The project made a moderate impact on the policy/policies in an instrumental way, or made an important contribution at a conceptual level to the policy/policies discussions.
 4 – The project made some identifiable impact on the policy/policies.
 2 – A claim for impact was made but no details given, or details given of a claim for expected future impacts.
 0 – No impact on policy making.

CHANGING HEALTH PRACTICE

We would like you to consider application of the findings by health practitioners on (inter)national or local level, such as the application of policies, changed behavior of practitioners and managers, the involvement of consumers/patients.

4a. Level of impact in health practice

10 – The project made an impact on behavior in more than 1 country.
 8 – The project made an impact on behavior at national level.
 6 – The project made an impact on the behavior of at least 1 team of practitioners or managers, or at least 1 group of patients/ members of the wider public.
 4 – The project made an impact on behavior of at least 1 or more practitioner, manager, patient, or member of the public.
 2 – A claim for impact made but no details given, or details given of a claim for expected future impacts.
 0 – The project made no impact on behavior.

4b. Degree of impact in health practice

10 – At least 1 of the changed behavior(s) was almost solely based on the project's evidence in a direct instrumental way.
 8 – The project made a considerable impact on the behavior(s).
 6 – The project made a moderate impact on the behavior(s).
 4 – The project made some identifiable impact on the behavior(s).
 2 – A claim for impact made but no details given, or details given of a claim for expected future impacts.
 0 – No impact on behavior.

BROADER IMPACT ON HEALTH

We would like you to consider the broader (societal) impact on health benefits, e.g., health gains, improved efficiency/cost reduction in the delivery of existing services, qualitative improvements in service delivery; improved allocation of resources at an area level, improved accessibility (e.g., improved geographical accessibility, improved equity of access on a gender or ethnic basis).

5a. Level of health benefits

10 – The project led to very large health benefits (in terms of a substantial health gain for many people, and/or major cost reductions, and so on).
 8 – The project led to considerable benefits (in terms of the amount of health gain, the numbers involved, and/or cost reductions, and so on).
 6 – The project led to moderate health benefits (in terms of the amount of health gain, the numbers involved, and/or cost reductions, and so on).
 4 – The project led to some health benefits (in terms of some health gain for some people, and/or some cost reduction, and so on).
 2 – A claim for impact made but no details given.
 0 – The project made no impact on health/health sector.

5b. Degree of impact on health benefits

10 – At least one of the health benefits was almost solely based on the project's evidence in a direct instrumental way.
 8 – The project made a considerable impact on the health benefits.
 6 – The project made a moderate impact on the health benefits.
 4 – The project made some identifiable impact on the health benefits.
 2 – A claim for impact made but no details given, or details given of a claim for expected future impacts.
 0 – No impact on health benefits.

Third Round Scoring

The detailed case studies were then sent to the HTA experts for re-scoring of the payback categories based on this fuller information now available. The experts were not provided with their initial scores, although they may have remembered them or recorded them themselves. These third-round scores were then compared with their individual second round scores. This comparison provided an indication of the influence of more detailed information on scoring of payback.

RESULTS

Output of HTA Projects

The findings show that the payback framework is a promising approach to assess the impact of HTA projects. We received thirty-four of forty-three surveys (response rate: 79 percent). In total, we identified over 101 peer reviewed papers; more than twenty-five PhD dissertations; citations of research in guidelines were mentioned in six projects; and implementation of new treatment strategies were described in eleven projects. The research payback of all thirty-four HTA projects is summarized (per subsidy round) in Table 2 below.

Scoring Impact

The HTA experts scored those HTA projects that had already led to a final report and for which we had received survey results ($n = 26$). All HTA experts used the payback categories and related scales (Table 1) at their disposal; they were not hesitant to score high or low (the average score given by the experts was 4.3). We could possibly, therefore, deduce that the descriptions of the scoring scales did not constrain the scoring range.

With regard to disagreement between the HTA experts in scoring, we conclude that whatever method of determining disagreement is used, most contention exists for the scoring items of scoring categories 3 to 5 (Informing Policy, Changing Health Practice, and Broader Impact on Health) (Table 3).

The participants at the expert meeting did not make much use of the possibility to re-score projects at the end of the meeting. Four out of nine participants changed their scores (for more than one study). In general, scores were adjusted to a lower score (dismissing claims of expected outcome), but also some higher scores were given. The re-scoring did reduce the variance around the scores for some projects. A preliminary conclusion is that discussing the projects in an expert meeting may lead to greater consensus. However, it would have been necessary to have discussed and rescored all projects, including those which initially had little variance, to be sure about this conclusion.

Analysis Based on the Scoring of the Impact of HTA Projects

Once the score (per payback category) per study has been calculated, a comparison can be made between the different subsidy rounds of the HTA program. For this, we averaged the scores per category per round (shown in Figure 2). This plot shows that, for all the five scoring categories, the average scores for each subsidy round remain quite close to each other.

Specific HTA projects can, however, potentially influence the scoring drastically. One clear example involves the PET subsidy round, which scored significantly less on both the Knowledge Production as well as the Research Benefits categories. Before attempting to explain this difference, we should realize that the average is based on small numbers; therefore, chance findings can easily occur. However, even when removing the project in this group that scored extremely low because of its early termination, the difference with the other projects remains.
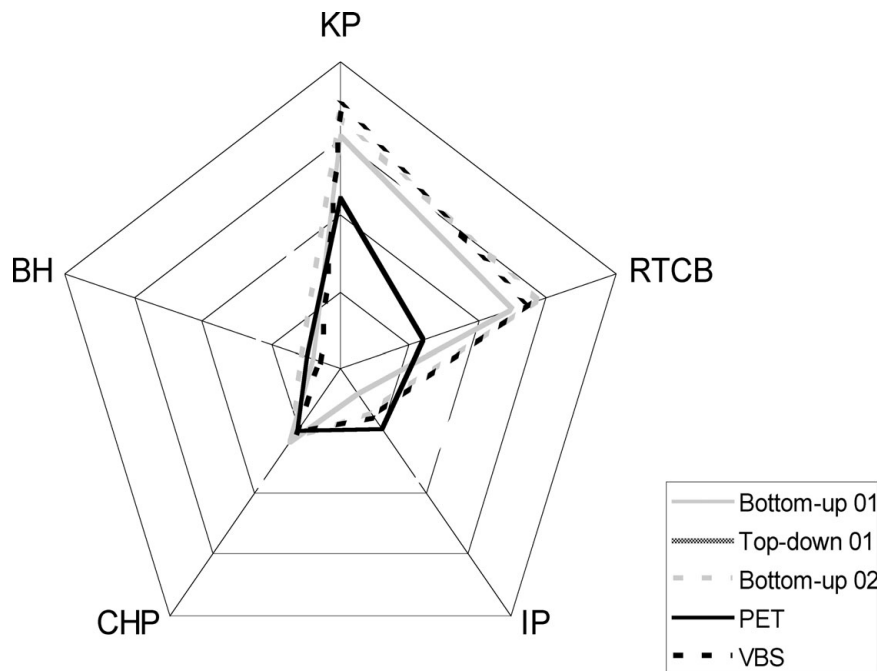

Figure 2. Average score per subsidy round, first scoring. KP, Knowledge Production; RTCB, Research Benefits; IP, Informing Policy; CHP, Changing Health Practice; BH, Broader Impact on Health.

Table 2. Summary of Research Payback of the Subsidy Rounds Considered

Subsidy round	Knowledge production	Research benefits	Informing policy	Changes in health practice	Broader impact on health
Bottom-up 2001 (N = 6/9 projects)	32 peer reviewed articles and 20 other publications (such as published abstracts, thesis and chapter in book). 75 presentations of which 21 international full presentations. The main audiences were academia and health professionals (e.g., psychiatrists and mental health workers).	Three projects resulted in 5 future projects of which one was a follow-on activity supported by ZonMw (total budget of 270,000 Euro). In total, 9 researchers obtained their PhD (resulting from 5 projects). Also, one project resulted in scientific training for medical students.	Informing actual policy was only mentioned by 2 projects. Considering the objectives of the DO program, informing policy is expected for almost all projects. However, there was little or no evidence of such impact available. One of the 6 DO projects has led to a revision of a guideline.	Overall impact on health practice was regarded as considerable in 4 projects (e.g., change of treatment strategies). In 5 projects considerable changes in health practice were expected, but often no evidence for expected changes was provided.	In 2 projects broader impact on health was regarded as moderate or considerable (e.g., reduced risk and patient burden for diagnostic imaging tests for peripheral arterial disease). Generally there was a lack of evidence for broader impact on health.
Top-down 2001 (N = 2/3 projects)	6 peer reviewed articles; 5 other publications (such as journal editorial, published abstract, and thesis). 18 presentations of which 7 international full presentations. The main audiences were academia and health professionals.	One of the projects led to further research that is performed in Belgium. In addition, the research team of this project appeared to be involved in writing a protocol for an international study. In one project a researcher has obtained his/her PhD in 2004. The other project is expected to provide a researcher with his/her PhD in 2007.	The results of both projects led to informing policy (e.g., by organizing international workshops on lower rectal cancer) and are expected to inform policy in the future (e.g., imaging of a rectal tumor before treatment will become mandatory and will be included in treatment protocols).	The results of one project led to the implementation of a new treatment strategy, which has become routine practice in several hospitals (i.e. use of less mediastinoscopies and thoracotomies in staging patients with lung cancer). The other project is expected to change primary treatment of rectal cancer patients by routinely scanning the liver.	Both projects expect to have an impact on health (e.g., improving the outcome of treatment for rectal cancer and less invasive techniques for staging of lung cancer will be more convenient for patients). However, no evidence on broader impact on health was provided.
Bottom-up 2002 (N = 13/18 projects)	33 peer reviewed articles; 42 other publications (such as published abstracts, newspaper articles and non-peer reviewed articles). 97 presentations of which 52 international full presentations. The main audiences were academia, health professionals and health users.	5 PhDs and one professorship have resulted and 12 PhDs are intended to result from the bottom-up projects. 13 follow-on projects totaling approximately 5 million Euro are related to these DO projects. Several follow-on activities (N = 4) are funded by ZonMw, but also by international funds such as the European Union. In addition, 2 projects resulted in contributions to (international) research teams.	2 projects have already been informing policy (e.g., a financial agreement with a local health insurer was made to introduce in vitro fertilization in manipulated natural cycles, development of guidelines for detection of hereditary colorectal cancer). Five other projects are expected to informing policy (e.g., affecting guidelines and hospital policy regarding spine fractures and an integrated diagnostic facility for psycho-geriatric patients).	In general, the results show that it is too early to tell whether the projects have changed practice. However, some examples presented were: change of IVF treatment practice in a few hospitals; and increase of referrals of children with anxiety disorders to individual cognitive behavioral therapy.	Generally there was a lack of evidence for broader impact on health.

Infertility disorders (VBS) ($N =$ 6/6 projects)	20 peer reviewed articles; 17 other publications (e.g., journal letters, conference proceedings, and non-peer-reviewed articles). 61 presentations of which 24 international full presentations. The main audience was academia and to a lesser extent health professionals.	3 projects contributed to 3 follow-on projects. ZonMw funded two of these projects (approximately 750.000 Euro in total). These projects have (partly) contributed to 4 PhDs and will (partly) result in 9 other PhDs during the coming 3 years. The research team of one project has also contributed to establishing a network of collaborative fertility physicians.	In general, the results show that it is too early to tell whether the projects have informed policy. In 2 projects it was mentioned that the results were used in revising an existing guideline and implementing a guideline. All projects are expecting to inform policy, e.g., by citing the findings in clinical guidelines, and by revising the current reimbursement system (per cycle).	In general, the results show that it is too early to tell whether the projects have changed practice. However, the results of 2 projects have been used in the implementation of a different treatment strategy in several hospitals. Some projects expected to change health practice, e.g., by limiting IVF treatment to couples with a clear diagnosis of infertility and by introduction of a policy of single embryo transfer.	Generally there was a lack of evidence for broader impact on health.
PET diagnostics ($N =$ 7/7 projects)	10 peer reviewed articles; 14 other publications (including conference proceedings, thesis, and published abstracts). 18 presentations of which 9 international full presentations. The main audience was academia.	5 PhDs resulted from 2 PET projects. In 2007, 1 PhD is expected to result from another project. 2 projects resulted in 3 follow-on projects, of which ZonMw funded one study (budget unknown).	The results of 2 projects are used in (drafting) guidelines (e.g., the guidelines for hypopharyngeal cancer). All projects are expecting to inform policy, e.g., by adaptation of guidelines for use of PET in staging of esophageal cancer or developing a guideline for use of PET for patients with Non-Hodgkin's lymphoma cancer.	There was a lack of evidence for changes in health practice. However, in 3 projects it was mentioned that the results were used in practice (in a protocol for detecting unknown primary tumors, in local guidelines for patients with head and neck cell carcinoma and local implementation of FDG-PET scan in patients with colorectal liver metastases). All projects expected to change health practice but often no evidence was provided.	Generally there was a lack of evidence for broader impact on health. The project on screening for distant metastases and synchronous primary tumors was an exception - in this project, the combination of CT/PET appeared to be more effective and cost saving compared with CT and PET separately.

Table 3. Different Measures of Disagreement for Each Payback Criterion

	1a	1b	2a	2b	3a	3b	4a	4b	5a	5b
Disagreement according to RAM	—	—	—	—	1 (4%)	3 (11%)	5 (19%)	3 (11%)	2 (8%)	3 (11%)
Disagreement according to 3 or more from median	6 (3%)	4 (2%)	8 (4%)	21 (11%)	33 (17%)	37 (19%)	31 (16%)	38 (19%)	34 (17%)	29 (15%)
Disagreement according to 2 or more from median	23 (12%)	22 (11%)	16 (8%)	45 (23%)	59 (30%)	77 (39%)	65 (33%)	79 (40%)	65 (33%)	66 (33%)

Note. The RAM score (row 1) is measured per project, so the number represents the disagreement out of 26 (projects). The other two scores (row 2 and 3) represent disagreement among 198 scores (for each of the 10 items, two experts scored only 19 relevant projects). 1a = Presentations; 1b = Publications (Knowledge Production); 2a = Capacity building; 2b = Research targeting (Research Benefits); 3a = Nature of the policy; 3b = Degree of impact on health policy (Informing Policy); 4a = Level of impact in health practice; 4b = Degree of impact in health practice (Changing Health Practice); 5a = Level of health benefits; 5b = Degree of impact on health benefits (Broader Impact on Health).

The case studies, although limited in number, showed that the more detailed analysis can add greater depth and understanding about the levels of impact that arise and why and how they have been achieved. Due to a low response rate in the third round (we received four to six scores per case study), the results are not presented here. However, the exercise illustrates a proof of concept that could be pursued when setting up a systematic research impact evaluation in future.

CONCLUSIONS

The payback framework is a comprehensive method to assess payback in a systematic way. In the payback framework, any assessment of the scientific quality of research (e.g., journal articles, the training of future researchers, and the development of careers) is part of the broader assessment of impact: the societal impact of research is the key issue in the multidimensional categorization of the benefits from health research (12).

The findings confirm that the payback framework is a useful approach to assess the impact of HTA projects. We were able to identify over 101 peer-reviewed papers, more than twenty-five PhD dissertations, citations of research in guidelines (according to six projects), and the implementation of new treatment strategies was mentioned in eleven projects. However, we found that assessing the actual impact of HTA projects too soon after the completion of the research is problematic and that assessing expected impacts is difficult. It is probably best to wait at least 3–5 years before addressing (policy) impact questions, although knowledge production and research benefits (publications, number of PhD dissertations, and so on) can be addressed earlier on. To track and document paybacks on a routine basis, it would be important to use a database that could be added to over time, and to return to principal investigators periodically after they had finished their research to update the information on paybacks.

On the basis of our findings, we conclude the following.

- In general, the payback framework is applicable to the Dutch HTA program. We were able to identify a considerable range of research paybacks. Given this diversity of outcome it is important to use an assessment method that captures a broad range of payback. The payback categories developed in this study provide a structured basis for any further monitoring or evaluation of outputs and outcomes of research funded by the HTA program.

- It is possible to score HTA projects, but the methodology used for scoring impact needs to be further developed. The results show that the methodology used for scoring impact should be refined—in particular the operationalization of the scoring scales, provision of sufficient information about the HTA projects and the scoring exercise itself. Communication to experts and coordination of experts' input is, therefore, a key issue. The potential of payback scoring has recently been promoted in an analysis of ways to evaluate research in a way that recognizes its value to users (5).

- Assessing and monitoring payback of related projects may be most appropriate. The experts of the scoring exercise believed that changes on the (inter)national level can generally not be attributed to single HTA projects. Pooling results of HTA projects in the same research area probably results in a higher impact compared with the impact of individual projects.

- This study was performed too early to fully assess the impact of the Dutch HTA program.

We recognized in advance that it would be difficult to obtain "hard" data for assessing the full impact due to the fact that the HTA program had only started in 2000. However, at the expert meeting, its value as a baseline measurement was underlined. Furthermore, the findings of this study can help funding organizations consider how to assess impact more

widely in the future, possibly routinely. This, in turn, could help inform research strategies and justify expenditure.

CONTACT INFORMATION

Wija J. Oortwijn, PhD (wija.oortwijn@ecorys.com), Senior Consultant, Department of Marco & Sector Policies, ECORYS Nederland BV, Watermanweg 44, 3067 GG, Rotterdam, The Natherlands

Stephen R. Hanney, PhD (stephen.hanney@brunel.ac.uk), Senior Research Fellow, Health Economics Research Group, Brunel University, Kingston Lane, Uxbridge, Middlesex, UB8 3PH, UK

Andreas Ligtvoet, MSc (a.ligtvoet@tudelft.nl), PhD researcher, Energy & Industry, Faculty of Technology, Policy and Management, Delft University of Technology, Jaffalaan 5, 2628 BX Delft, the Netherlands

Stijn Hoorens, MSc (hoorens@rand.org), Senior Analyst, **Steven Wooding**, PhD (wooding@rand.org), Team Leader, Science and Technology, **Jonathan Grant**, PhD (jgrant@rand.org), President of RAND Europe Cambridge Ltd., RAND Europe Cambridge Ltd., Westbrook Centre, Milton Road, Cambridge, CB4 1YG

Martin J Buxton, BA (Soc Sci) (martin.buxton@brunel.ac.uk), Professor of Health Economics, Health Economics Research Group, Brunel University, Kingston Lane, Uxbridge, Middlesex, UB8 3PH, UK

Lex M. Bouter, PhD (lm.bouter@dienst.vu.nl), Rector, Executive Board, VU University Amsterdam, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands

REFERENCES

1. Bouter LM, Knottnerus JA. Beoordeling van maatschappelijke relevantie van toegepast gezondheidsonderzoek: het belang van publiceren in nationale vaktijdschriften als ruwe indicator [Assessing the societal relevance of applied health research: the use of publications in national journals as an indicator]. *Ned Tijdschr voor Geneeskd*. 2000;144:1178-1183.
2. Buxton M, Hanney S. *Assessing payback from Department of Health Research & Development*. vol. 1. The main report. Research report #19. Uxbridge: HERG, Brunel University; 1994.
3. Buxton M, Hanney S. How can payback from health services research be assessed? *J Health Serv Res Policy*. 1996;1:35-43.
4. Buxton M, Hanney S, Packwood T, Roberts S, Youll P. Assessing benefits from Department of Health and National Health Service R&D. *Public Money Manage.* 2000;20:29-34.
5. Donovan C. The qualitative future of research evaluation. *Sci Public Policy*. 2007;34:585-597.
6. EMGO. Institute for Research in Extramural Medicine. *Annual Report 2006. Chap 5. "Societal Impact."* Amsterdam: EMGO Institute; 2007.
7. Fitch K, Bernstein SJ, Dolores Aguilar M, et al. *The RAND/UCLA Appropriateness Method user's manual*, MR-1269-DG-XII/RE. Santa Monica: RAND Corporation; 2001.
8. Hanney S, Buxton M, Green C, Coulson D, Raftery J. An assessment of the impact of the NHS Health Technology Assessment Programme. *Int J Health Technol Assess*. 2007;11.
9. Hanney S, Soper B, Buxton M. *Evaluation of the NHS R&D Implementation Methods Programme. HERG Research Report No 29*. Uxbridge: HERG, Brunel University; 2003.
10. Hanney S, Grant J, Wooding S, Buxton M. Proposed methods for reviewing the outcomes of health research: the impact of funding by the UK's 'Arthritis Research Campaign.' *Health Res Policy Syst*. 2004;2:4.
11. National Institutes of Health. *Research program outcomes assessment material*. Bethesda, MD: National Institutes of Health; 2000.
12. Raad voor Gezondheidsonderzoek. Onderzoek dat ertoe doet. *De responsiviteit van universitair medische centra op vraagstukken in volksgezondheid en gezondheidszorg*. No. 57 [*Research that matters. Responsiveness of university hospital regarding public health and health care issues*]. Den Haag: Raad voor Gezondheidsonderzoek; 2007.
13. Royal Netherlands Academy of Arts and Sciences. *The societal impact of applied research towards a quality assessment system*. Amsterdam: Royal Netherlands Academy of Arts and Sciences; 2002.
14. Smith R. Measuring the social impact of research. *BMJ*. 2001;323:528.
15. Van Weel C. Biomedical science matters for people—so its impact should be better assessed. *Lancet*. 2001;360:1034-1035.
16. Wooding S, Hanney S, Buxton M, Grant J. Payback arising from research funding: Evaluation of the Arthritis Research Campaign. *Rheumatology*. 2005;44:1145-1156.
17. ZonMw. *Annual Report 2001 of the ZonMw programme health care efficiency research*. The Hague: ZonMw; 2002.
18. ZonMw. *Annual Report 2002 of the ZonMw programme health care efficiency research*. The Hague: ZonMw; 2003.