



EUROPE

CHILDREN AND FAMILIES
EDUCATION AND THE ARTS
ENERGY AND ENVIRONMENT
HEALTH AND HEALTH CARE
INFRASTRUCTURE AND
TRANSPORTATION
INTERNATIONAL AFFAIRS
LAW AND BUSINESS
NATIONAL SECURITY
POPULATION AND AGING
PUBLIC SAFETY
SCIENCE AND TECHNOLOGY
TERRORISM AND
HOMELAND SECURITY

The RAND Corporation is a nonprofit institution that helps improve policy and decisionmaking through research and analysis.

This electronic document was made available from www.rand.org as a public service of the RAND Corporation.

Skip all front matter: [Jump to Page 1](#) ▼

Support RAND

[Browse Reports & Bookstore](#)

[Make a charitable contribution](#)

For More Information

Visit RAND at www.rand.org

Explore [RAND Europe](#)

View [document details](#)

Limited Electronic Distribution Rights

This document and trademark(s) contained herein are protected by law as indicated in a notice appearing later in this work. This electronic representation of RAND intellectual property is provided for non-commercial use only. Unauthorized posting of RAND electronic documents to a non-RAND Web site is prohibited. RAND electronic documents are protected under copyright law. Permission is required from RAND to reproduce, or reuse in another form, any of our research documents for commercial use. For information on reprint and linking permissions, please see [RAND Permissions](#).

This product is part of the RAND Corporation technical report series. Reports may include research findings on a specific topic that is limited in scope; present discussions of the methodology employed in research; provide literature reviews, survey instruments, modeling exercises, guidelines for practitioners and research professionals, and supporting documentation; or deliver preliminary findings. All RAND reports undergo rigorous peer review to ensure that they meet high standards for research quality and objectivity.

R E P O R T



Project Retrosight

Understanding the returns from
cardiovascular and stroke research

Methodology Report

Alexandra Pollitt, Steven Wooding,
Stephen Hanney, Martin Buxton, Jonathan Grant

Supported by the National Institute for Health Research,
the Canadian Institutes of Health Research, the Heart and Stroke
Foundation of Canada, and the National Heart Foundation of Australia

The research described in this report was supported by the National Institute for Health Research, the Canadian Institutes of Health Research, the Heart and Stroke Foundation of Canada, and the National Heart Foundation of Australia.

RAND Europe is an independent, not-for-profit research organisation whose mission is to help improve policy and decision making through research and analysis. RAND's publications do not necessarily reflect the opinions of its research clients and sponsors.

RAND® is a registered trademark.

© Copyright 2011 RAND Corporation

Permission is given to duplicate this document for personal use only, as long as it is unaltered and complete. Copies may not be duplicated for commercial purposes. Unauthorized posting of RAND documents to a non-RAND website is prohibited. RAND documents are protected under copyright law. For information on reprint and linking permissions, please visit the RAND permissions page (<http://www.rand.org/publications/permissions.html>).

Published 2011 by the RAND Corporation
1776 Main Street, P.O. Box 2138, Santa Monica, CA 90407-2138
1200 South Hayes Street, Arlington, VA 22202-5050
4570 Fifth Avenue, Suite 600, Pittsburgh, PA 15213-2665
Westbrook Centre, Milton Road, Cambridge CB4 1YG, United Kingdom
RAND URL: <http://www.rand.org>
RAND Europe URL: <http://www.rand.org/randeurope>
To order RAND documents or to obtain additional information, contact
Distribution Services: Telephone: (310) 451-7002;
Fax: (310) 451-6915; Email: order@rand.org

Preface

This report describes in detail the methodology for a case study-based review of 29 cardiovascular and stroke research grants, funded in three countries between 1989 and 1993. The case studies focused on the individual grants but considered the development of the investigators and ideas involved in the research projects from initiation to the present day. Basic biomedical and clinical cardiovascular and stroke research grants awarded in Australia, Canada and the UK were selected through a stratified random selection approach that aimed to include both high- and low-impact grants. The structured case studies were used to assess the impact of these grants in detail and to draw conclusions about the relationship between impact and a range of factors. The key messages from the project are as follows.

- The cases reveal that a large and diverse range of impacts arose from the 29 grants studied.
- There are variations between the impacts derived from basic biomedical and clinical research.
- There is no correlation between knowledge production and wider impacts.
- The majority of economic impacts identified come from a minority of projects.
- We identified factors that appear to be associated with high and low impact.

This report presents the project's methodology in detail. The Policy Report¹ presents the project's findings and their implications for policy, and the Case Study report² presents all 29 case studies carried out.

This work was led by RAND Europe in collaboration with the Health Economics Research Group (HERG) at Brunel University. RAND Europe is an independent not-for-profit policy research organisation that serves the public interest by improving policy-making and informing public debate. The Health Economics Research Group is a Specialist Research Institute of Brunel University dedicated to conducting accessible, policy-relevant research of a high academic quality focused on improving the efficiency and cost-effectiveness of resources devoted to health care and to research. This report has been

¹ Wooding, S., S. Hanney, A. Pollitt, M. Buxton, and J. Grant. *Project Retrosight. Understanding the returns from cardiovascular and stroke research: Policy Report*. Cambridge, UK: RAND Europe, MG-1079-RS, 2011.

² Pollitt, A., S. Wooding, S. Hanney, M. Buxton, and J. Grant. *Project Retrosight. Understanding the returns from cardiovascular and stroke research: Case Studies*. Cambridge, UK: RAND Europe, WR-836-RS, 2011.

peer-reviewed in accordance with RAND's quality assurance standards³ and therefore may be represented as a RAND Europe product.

For more information about RAND Europe or this document, please contact:

Steven Wooding
Research Leader

RAND Europe
Westbrook Centre, Milton Road
Cambridge, CB4 1YG
United Kingdom

Email: wooding@rand.org or reinfo@rand.org

³ RAND Europe: see <http://www.rand.org/about/standards/> [Accessed 16 December 2010]

Contents

Preface.....	i
Table of figures.....	v
Table of tables.....	vii
Acknowledgements and detailed authorship.....	ix
Consortium members.....	xi
CHAPTER 1 Introduction.....	1
1.1 Summary of methods.....	2
CHAPTER 2 Framing issues.....	5
2.1 Introduction.....	5
2.2 Project origins.....	5
2.3 The case study approach.....	6
2.4 The Payback Framework.....	7
2.4.1 Payback classification.....	7
2.4.2 The logic model.....	9
2.5 Unit of analysis.....	12
2.6 Selection of time frame.....	13
2.7 Selection of characteristics of interest and the selection matrix.....	13
2.8 Factors used for selecting case studies.....	13
2.9 The selection matrix.....	16
CHAPTER 3 Identifying candidates for case studies.....	17
3.1 Introduction.....	17
3.2 Grant identification.....	17
3.3 PI survey.....	18
3.3.1 Survey outline.....	18
3.3.2 Building the survey tool.....	19
3.3.3 Testing the survey tool.....	19
3.3.4 Survey deployment and response rates.....	19
3.3.5 Identifying email addresses.....	20
CHAPTER 4 Case study selection.....	21
4.1 Filling the matrix.....	21

4.2	Selecting the cases.....	22
CHAPTER 5 Case study research.....		23
5.1	Pilot case study.....	23
5.2	Cross-country working.....	24
5.3	Case study research methods.....	25
5.3.1	Document and literature review.....	25
5.3.2	Interviews.....	25
5.3.3	Bibliometric analysis.....	26
5.4	Case study write-up and review.....	30
5.5	External review.....	30
CHAPTER 6 Rating impact.....		33
6.1	The rating panel.....	33
6.2	Rating scale.....	34
6.3	Rating material.....	35
6.4	Rating process.....	36
6.5	Initial rating.....	36
6.6	Workshop.....	37
CHAPTER 7 Data characterisation and robustness testing.....		39
7.1	Potential sample bias.....	39
7.2	Rating behaviour.....	41
7.3	Confounder analysis.....	45
7.3.1	Clinical and basic biomedical distinction.....	45
7.3.2	Case study scope.....	46
7.3.3	Negative or null findings.....	47
CHAPTER 8 Analysis.....		49
8.1	Deriving impact categories.....	49
8.2	Identifying factors associated with research impact.....	52
8.3	Emerging Findings workshop.....	55
REFERENCES.....		57
APPENDICES.....		61
	Appendix A: PI Survey.....	63
	Appendix B: PI interview protocol.....	67
	Appendix C: Non-PI interview protocol.....	75
	Appendix D: External peer review materials.....	83
	Appendix E: Rating instructions.....	89

Table of figures

Figure 1-1: A diagrammatic summary of the project methods	3
Figure 2-1: The Payback Framework’s logic model of the research funding process.....	9
Figure 2-2: Selection matrix.....	16
Figure 5-1: Guidelines for paper attribution.....	27
Figure 5-2: Details of analysis methods and indicators used	29
Figure 6-1: Illustrative section of rating sheet.....	37
Figure 7-1: Total number of years the PIs responding to our survey had spent in research by 2007	40
Figure 7-2: Distribution of exclusions: completed case studies are shown as filled circles, refusals and exclusions as open circles	40
Figure 7-3: Ratings distributions (by person and category)	42
Figure 7-4: Ratings agreement	44
Figure 7-5: Average deviation from the median before and after the workshop.....	45
Figure 7-6: A comparison of the alternative research type definitions as applied to the case studies.....	46
Figure 8-1: Payback categories and impact groupings.....	50
Figure 8-2: A diagrammatic overview of the analysis process to identify factors associated with impact.....	51

Table of tables

Table 3-1: The number of PIs, email addresses and level of bibliometric activity.....	19
Table 3-2: The number of PIs contacted and the number who responded in each country	20
Table 4-1: Number of grants we received survey responses on in each cell of the selection matrix.....	21
Table 6-1: Rating panel membership	33
Table 7-1: Summary of the mean and median ratings pre- and post- workshop	43
Table 8-1: The correlations between ratings in each payback category.....	49
Table 8-2: Overlap between high, mid and low groupings for academic and wider impact.....	52
Table 8-3: Overlap between the high, mid and low groups for all cases studies versus the groupings for basic and clinical case studies separately.....	52
Table 8-4: Emerging Findings workshop participants	55

Acknowledgements and detailed authorship

We would like to start by acknowledging all the scientists who were willing to act as the participants for this study, particularly the principal investigators of the 29 case study grants. The study would clearly have been impossible without them.

We also owe a debt of gratitude to the external experts who participated in our rating workshop: Cy Frank, Liz Allen, Brendan Curran and Marcus Nichol. We would also like to thank those scientists who provided reviews of each of our case studies, and participated in our Emerging Findings Workshop in April 2010 in London.

This study was initiated with internal funding from RAND Europe and HERG, with continuing funding from the UK National Institute for Health Research, the Canadian Institutes of Health Research, the Heart and Stroke Foundation of Canada and the National Heart Foundation of Australia. The UK Stroke Association and the British Heart Foundation provided support in kind through access to their archives.

Project Retrosight team

Jonathan Grant (RAND Europe), Martin Buxton (HERG), Stephen Hanney (HERG) and Steven Wooding (RAND Europe) devised the methodological approach and analysis with input from the Retrosight team. Steven Wooding managed and coordinated the project, supported by Eddy Nason (RAND Europe), Sharif Ismail (RAND Europe), Sue Kirk (RAND Europe) and Alex Pollitt (RAND Europe). Statistical analysis was provided by Laura Staetsksy (RAND Europe).

The Australian case studies were carried out by Rob Mercer (Instinct and Reason), Angela Mitchell (Instinct and Reason) and Christine Latif (National Heart Foundation of Australia); the Canadian studies by Laura McAuley (Canadian Institutes of Health Research), Heather Mustoe (Canadian Institutes of Health Research) and Kimberly-Anne Ford (Canadian Institutes of Health Research); the UK studies by Stephen Hanney, Sharif Ismail, Sue Kirk, Sonja Marjanovic (RAND Europe) and Eddy Nason. The case study bibliometric analysis was coordinated by Linda Butler (Australian National University) and Sharif Ismail and carried out by Kumara Henadeera (Australian National University) and Thed van Leeuwen (CWTS).

Quality assurance

Tom Ling (RAND Europe) and Chris Henshall (Brunel University and the University of York) acted as our quality assurance reviewers and provided thoughtful, constructive and timely criticism.

Steering group

The project steering group consisted of Peggy Borbey (Canadian Institutes of Health Research), Sally Brown (Heart and Stroke Foundation of Canada), Martin Buxton, Ian Graham (Canadian Institutes of Health Research), Jonathan Grant, Stephen Hanney, Christine Latif, Laura McAuley and Linda Piazza (Heart and Stroke Foundation of Canada).

Report writing

The report was drafted by David C. Taylor (Cambridge Editorial) working with the RAND and HERG project teams and edited by Sue Ecob (Cambridge Editorial).

Five babies were born to members of the project team during Project Retrosight: Noah Grant Sanz, Alex Robert Wooding, Kaya Diane Hipwell, Lena Grae Simpson, and Daniel Peter Wooding.

Consortium members

A large number of organisations supported Project Retrosight.

RAND Europe

RAND Europe is an independent not-for-profit policy research organisation that helps improve policy and decision making through research and analysis. Its clients are European governments, institutions, foundations and companies with a need for rigorous, impartial, multidisciplinary analysis.

HERG

The Health Economics Research Group is a Specialist Research Institute of Brunel University dedicated to conducting accessible, policy-relevant research of a high academic quality focused on improving the efficiency and cost-effectiveness of resources devoted to health care and to research.

NIHR

The goal of the National Institute for Health Research (NIHR) is to create a health research system in which the National Health Service (NHS) supports outstanding individuals, working in world-class facilities, to conduct cutting-edge research focused on the needs of patients and the public. The NHS's reputation for international excellence is growing as it gains recognition for being the preferred host for collaborative and multicentre research in the public interest in partnership with and for industry.

CIHR

The Canadian Institutes of Health Research (CIHR) is the Canadian government's agency responsible for funding health research. CIHR was created in 2000 under the authority of the CIHR Act and reports to parliament through the Minister of Health. CIHR is comprised of 13 institutes and supports 13,000 researchers and research trainees. CIHR's mandate is to "excel, according to internationally accepted standards of scientific excellence, in the creation of new knowledge and its translation into improved health for Canadians, more effective health services and products and a strengthened Canadian health-care system".

HSFC

The Heart and Stroke Foundation of Canada is a federation of ten provincial foundations with a federation office in Ottawa, led and supported by more than 130,000 volunteers. The mission of HSFC, a volunteer-based health charity, is to lead in eliminating heart disease and stroke and reduce their impact through the advancement and application of research and the promotion of healthy living and advocacy.

National Heart Foundation of Australia

The National Heart Foundation of Australia is a non-profit organisation with a mission “to reduce suffering and death from heart, stroke and blood vessel disease in Australia”. Established in 1959, the National Heart Foundation funds world-class cardiovascular research, supports health professionals in their practice, develops health promotion activities, informs and educates the public and assists people with cardiovascular disease.

Instinct and Reason

Instinct and Reason is a global research consultancy that works with private, public sector (state and federal) and not-for-profit organisations. Instinct and Reason specialises in both domestic and international market, social and economic research and has offices in Canberra, Sydney, New York and London.

The Research Evaluation and Performance Project

The Research Evaluation and Performance Project was a research unit of the Australian National University, focusing on the development of methods to assess the impact of research, with a particular expertise in bibliometric methods.

The Centre for Science & Technology Studies

The Centre for Science & Technology Studies (CWTS) is a research institute within the Leiden University Faculty of Social Sciences. CWTS is a leading research institute in the field of scientometrics and bibliometrics, whose bibliometric techniques are frequently used in research assessment procedures. CWTS’s clients include international organisations, governments and research groups both inside and outside the Netherlands.

Stroke Association

The Stroke Association campaigns, educates and informs to increase knowledge of stroke at all levels of society, acting as a voice for everyone affected by stroke. The charity funds research into prevention, treatment and better methods of rehabilitation; it helps stroke patients and their families directly through its community support services as well as providing information through its helpline, leaflets and factsheets.

British Heart Foundation

The British Heart Foundation is the nation’s heart charity, dedicated to saving lives through pioneering research, patient care, campaigning for change and by providing vital information. It works alongside government, other health charities, health professionals and thousands of dedicated supporters to beat heart disease.

Project Retrosight was a multinational study that investigated the translation and payback from basic biomedical and clinical cardiovascular and stroke research projects. The main project aims were to:

- examine the variety of payback produced by basic biomedical and clinical cardiovascular and stroke research;
- identify factors associated with high (and low) levels of payback, in particular factors relating to the characteristics of the research, how it was supported or the context in which it was carried out;
- examine the significance of country context in the process of translation and production of payback. In the event it did not prove possible to carry out cross-country analysis (discussed in the Policy Report, Wooding et al., 2011).

These basic aims informed all methodological decisions made during the project. This report describes the methodology in detail, working through the project stage by stage.

The resulting analysis is based on a series of 29 case studies of basic biomedical and clinical cardiovascular and stroke research project grants in Australia, Canada and the UK. Each case study focuses on a single project grant awarded to a researcher during the selected time window of 1989–93 and follows the process that led from the application for the grant and its subsequent award, through the research that followed and the outputs emanating from it up to the present day. A standard case study selection process was used for each grant in each of the three countries.

Our analysis compared the payback for each grant and enabled us to draw out common factors associated with impact. The five key findings, which are explored in detail in the Policy Report (Wooding et al., 2011), are as follows.

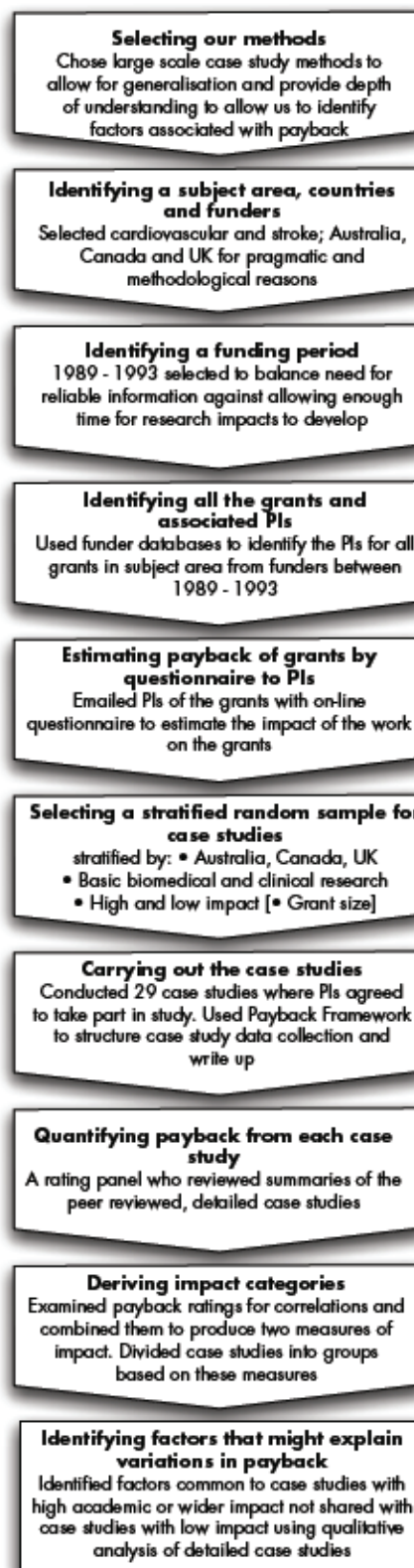
1. The cases reveal that a large and diverse range of impacts arose from the 29 grants studied.
2. There are variations between the impacts derived from basic biomedical and clinical research.
3. There is no correlation between knowledge production and wider impacts.

4. The majority of economic impacts identified come from a minority of projects.
5. We identified factors that appear to be associated with high and low impact.

1.1 **Summary of methods**

A summary of the Project Retrosight approach is presented in Figure 1.1. In this report we discuss each stage in turn, starting with how we selected our methods and identified the body of research to examine. We then step through the stages of our approach: collating lists of research grants for possible case study; estimating the payback of each grant through a survey of principal investigators (PIs); selecting a random stratified sample of grants for case study; building those case studies; quantifying the payback in each; and finally identifying common factors associated with payback.

Figure 1-1: A diagrammatic summary of the project methods



2.1 **Introduction**

Before beginning our research, we had to resolve a number of “framing issues” – fundamental decisions that would define the parameters of our study. In this section we describe these framing issues, explaining:

- the origins of the project;
- the decision to adopt a case study approach;
- the origin and application of the Payback Framework;
- why we chose the 1989–93 time window;
- how we decided which aspects of research to concentrate on.

2.2 **Project origins**

The project was inspired by series of conversations between Martin Buxton of the Health Economics Research Group at Brunel University (HERG) and Jonathan Grant of RAND Europe during breaks in a 1999 workshop on “research impact” organised in Banff, Canada, by the Alberta Heritage Foundation for Medical Research (now Alberta Innovates - Health Solutions). Interest in carrying out a study in the cardiovascular field in Canada was sparked by a similar study conducted by the project team in arthritis (Wooding et al., 2004, 2005), the results of which were presented by Steven Wooding at an event hosted by the Canadian Institutes of Health Research (CIHR) and attended by Sally Brown of the Heart and Stroke Foundation of Canada.

Assembling a consortium to support this multi-country study was time-consuming, and it was not until 2007 that work finally started on the project, as a joint undertaking between HERG and RAND Europe, with support from the CIHR, the Heart and Stroke Foundation of Canada, the National Heart Foundation of Australia, the UK National Institute of Health Research, the UK Stroke Association and the British Heart Foundation.

The project was managed by RAND Europe and UK case studies were carried out by researchers from both RAND and HERG. Canadian case studies were carried out by an in-house team of researchers from the CIHR while Australian case studies were conducted by a team comprising an internal staff member of the National Heart Foundation of Australia and a specialist consultancy, Instinct and Reason.

Bibliometric analysis for the project was coordinated by Linda Butler of the Research Evaluation and Performance Project at the Australian National University and data analysis was carried out in collaboration with Thed van Leeuwen of the University of Leiden's Centre for Science and Technology Studies (CWTS).

2.3 The case study approach

We chose the case study approach because we wanted to understand each grant and its story in detail. As with all research methods, this approach has its advantages and disadvantages. In general, case studies are preferred when “how” or “why” questions are being proposed as they provide a rich source of material for examination (Yin, 2003). In the context of Project Retrosight, they provide a detailed picture of what led to the award of the grant, how the research progressed and how it subsequently developed. Indeed, there is a long history of the use of case studies to examine the translation of research into actual results, especially where the emphasis of a study is on showing the long-term benefits from health research (Hanney et al., 2004).

For Project Retrosight the use of case studies enabled us to understand and demonstrate, in narrative form, how the output from research funded in the early 1990s was – or was not – translated into practical measures that have had an impact on healthcare and society at large.

Each case study provides an overview of both the long-term and shorter-term outcomes of a specific piece of work. Case studies are especially useful in illuminating research translation because they allow detailed exploration of contextual and individual factors that are (or may be) significant, and that are not easily captured by macro-level analysis.

When used for such illustrative purposes, case studies will not be representative of the wider research carried out at the time in a statistical sense; this would require a fully randomised, large-scale sample, which is simply not feasible when using an in-depth case study approach. Instead we have aimed to select a balance of case studies that encompasses the range of research taking place at the time. This was based on a number of key criteria; for example, we selected both basic biomedical and clinical research case studies in order to explore and compare the impact of these two types of research. The importance of achieving an appropriate balance meant that selecting suitable case studies was a critical stage of the research process.

Case studies are also effective when trying to build up a picture of typical behaviour. In this instance, we aimed to use the case study approach to provide insights into the research translation process that could help to improve the management and funding of research. Because of this it was important that the conclusions from the case studies could be generalised.

For case study research the justification for generalisation is the repetition of a finding across case studies rather than statistical representativeness. A finding seen in a substantial number of case studies can be more confidently generalised than one found in only one or two examples. In order to increase the power of such comparisons, the case studies on either side of the comparison may be selected to balance other characteristics. So, for

example, in a comparison of the effect of gender, comparison cases might be matched for the career position of the investigator.

In contrast, survey-based studies achieve generalisability through ensuring a statistically representative sample of the population. However, the depth in which individual cases can be examined, as well as the contextual information that can be incorporated, is much more limited.

The relatively large number of case studies in Project Retrosight gave us considerable insight and provided sufficiently robust evidence to make a variety of comparisons. In a study that adopts a multicase approach, the aim is to ensure that we can examine the benefits from different types of research and look for common factors. Using multiple case studies also allows us to carry out cross-case analysis from which to draw some general conclusions.

2.4 The Payback Framework

The research evaluation framework used in Project Retrosight was developed by HERG at Brunel University (Buxton and Hanney, 1996, 1997) and subsequently refined in collaboration with RAND Europe (Hanney et al., 2004). We chose to use it for this project because it is a proven framework for the selection and analysis of case studies in research evaluation. The Payback Framework is both a tool for evaluating a range of potential *outputs* from research and additionally (unlike most other frameworks) a logic model that provides a mechanism for conceptualising the *process* through which outputs are created.

The Framework consists of two components. The first is the system of classifying the outputs of research. It should be emphasised that the classification system is designed to cope with both quantifiable outcomes – for example, the number of research publications – and non-quantifiable impacts such as a description of career progression following the award of a research grant. The second component is a logic model of the research process. Logic models are widely used in evaluation methodology to understand input–process–output relationships and to break down research programmes into their constituent parts.

Originally designed for health service research in the UK, the Framework has since been applied in a number of contexts to evaluate payback from the funding of biomedical research by a Canadian organisation (Buxton and Schneider, 1999) and the analysis of health services research by the World Health Organisation (Pang et al., 2003). More recently it has been used to explore the payback from arthritis research funded by the Arthritis Research Council (Wooding et al., 2004) and to analyse the payback from health technology assessment programmes in the UK (Hanney et al., 2007) and the Netherlands (Oortwijn et al., 2008). The Framework has also been used to good effect in assessing social science, arts and humanities research (Wooding et al., 2007).

2.4.1 Payback classification

The evaluation framework considers five categories of impact: knowledge production, research targeting and capacity building, informing policy and product development, health and health sector benefits and broader economic benefits (described in detail below). A range of methods can be used to assess these individual research output

categories, including expert review, bibliometric analysis, social science methods and economic analysis.

Knowledge production

The knowledge produced by research is manifested primarily through a variety of publications. Any type of publication can be considered, but peer-reviewed articles are generally thought to be the most valuable and objective measure. Other published material includes editorials, meeting abstracts, reviews and patent filings. Citation analysis is a useful tool that is often used in assessing the impact on the research field of knowledge emanating from a research project.

Research targeting and capacity building

Both these components capture benefits for future research.

Research targeting

The knowledge produced by a research project often influences subsequent research, research agendas and funding priorities, both within a specific research area and in complementary fields. This influence can sometimes be identified through citation analysis. Information about funding sources and grant sizes for a specific publication (often indicated in the acknowledgements section of published material) can also help win support for follow-on research.

Capacity building

A research project can contribute to building capacity within the research organisation, in terms of both infrastructure and staff development. For example, a project may attract grant support for education, training and career development. This can be manifested in the number of higher degrees or professional promotions awarded, the type of research being funded and the publication of research papers by junior researchers. Improvements to infrastructure might include expanding or upgrading laboratory and office facilities and investment in technology and equipment.

Informing policy and product development

Research can be used to inform policy making in a wide range of circumstances and at a number of levels. The impact of a piece of research on health policy making can be measured through interviews with key personnel and through bibliometric indicators such as the appearance of research citations in clinical guidelines and other policy documents. Research can also contribute to technological and product development, with findings being fed into the private sector for commercialisation. This is often achieved by licensing intellectual property rights or contract research work, forming public-private joint ventures, and through new business start-ups.

Health and health sector benefits

The principal payback that most health research funders aim to achieve through their funded research is a discernible health benefit for their target population. These benefits materialise through the adoption by public and private sector bodies of the policy insights, products, practices or technologies emerging from the research.

In addition to direct improvements in patient care, the health sector itself can also benefit from research findings, for example, through cost savings and quality gains in service

provision, and these savings can be re-invested in improved health care provision and further research.

Broader economic benefits

The scientific, clinical and commercial application of a research project’s findings can deliver a wide range of socioeconomic benefits.

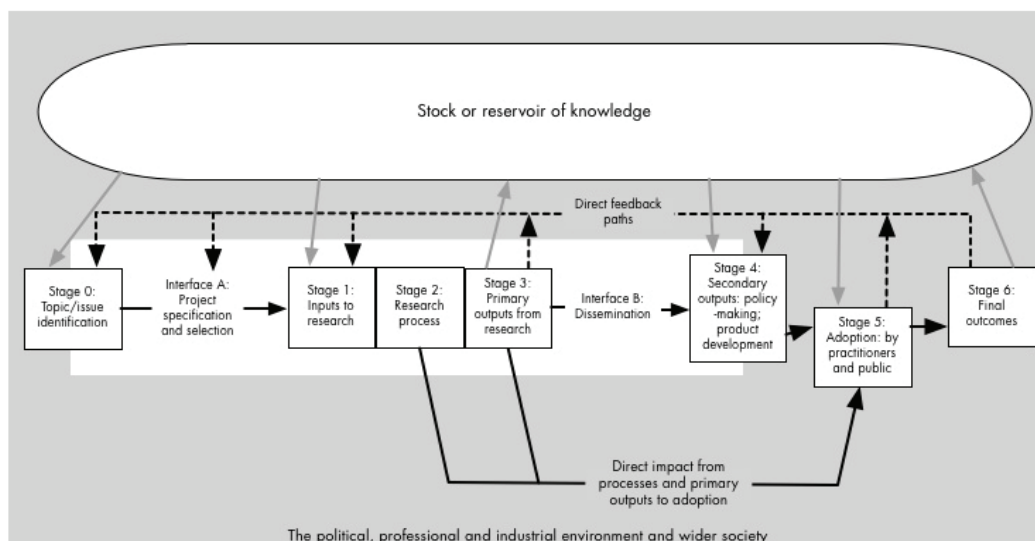
Exploitation of the intellectual property created by research can result in new employment opportunities and export potential, and may attract international investment in the national research base. The health and health sector benefits that accrue from health research projects contribute to the productivity of a healthy workforce and bring associated economic benefits. This type of impact is particularly challenging to measure, due to the difficulty of accurately attributing benefits to specific individual research projects.

It should be borne in mind that these five “payback categories” may apply in varying degrees according to the type of research. For example, when evaluating the outcomes from basic science research, it may be the case that knowledge production outputs are more common than outputs that inform policy. On the other hand, the outputs from a clinical research project might be expected to have a more direct policy influence. Evaluating the contribution of basic research to more downstream payback categories, such as policy formulation and broader economic benefits, might require, at the very least, a longer period of study than a similar evaluation of clinical research.

2.4.2 **The logic model**

The second element of the evaluation framework is the logic model, which describes the various stages in the process through which research can generate impacts (Figure 2.1). The model does not pretend to be a precise illustration of how research utilisation occurs; the processes through which research is translated into actual results are likely to be more complex than the model can present. Nevertheless, it provides a useful structure around which to organise the assessment of research impact over time and to enable comparisons between cases.

Figure 2-1: The Payback Framework’s logic model of the research funding process



Source: Hanney et al., 2004

There are seven stages in the logic model, from the initial idea for the research project through to the wider impact on society, and each of these is described in detail below.

Stage 0: Topic/issue identification

The assessment begins by investigating how the idea for the research was born. There are numerous potential sources of the original idea, and in many cases several of these contribute to identifying the research topic. They might include:

- an individual scientist's intellectual curiosity;
- an existing need (known within the research community) to fill certain gaps in scientific knowledge;
- a need identified as a result of personal experience (for example, a clinical researcher's personal experience in treating patients);
- opportunistic motives, such as availability of funding;
- externally solicited research in the form of requests for proposals by a research funding body or other interested stakeholder groups.

Interface A: Project specification and selection

Different funders exert different degrees of direction with respect to their research funding. In pure response mode funding – i.e. when the topic is identified by the research team – project specification and selection interface generally involve traditional processes of developing a detailed proposal and submitting it for peer review, possibly modifying it post-feedback, and then applying for funding (which is either granted or refused). In more directed funding, the peer-review process can involve a range of stakeholders and, in some cases, a competitive bidding process.

Stage 1: Inputs into research

This stage concerns the resources available to the project. These will include financial support, human resources (such as scientific expertise and administrative support), physical resources (scientific equipment, facilities and general consumables), collaborators and so on.

Some consideration of these inputs by the researchers will already have been required at Interface A, when the research proposal and budget requirements were drafted. Now, at Stage 1, these requirements may be reviewed and complementary inputs, in addition to those budgeted for by the project funding, can be considered.

These complementary inputs might include the experience and accumulated knowledge-base of the research team and the availability of equipment and research infrastructure, such as university laboratory and office space, administrative support and collaborator networks that can be tapped into. All these can influence the outputs from the research.

Stage 2: Process

Stage 2 evaluates the research process by exploring a number of key issues, including the suitability of the research design and methods for answering the central scientific question; the difficulties encountered in the research process; factors that facilitate or impede the research process; the time and cost efficiency of the research; the nature of collaborations;

interactions with potential users of the research as the project is undertaken; and any early dissemination of knowledge or adoption activities occurring as findings emerge incrementally.

The sort of factors that can facilitate or impede research may include the level of scientific challenge inherent in the programme; general resource availability (human, financial or physical); internal team dynamics, which can affect motivation and productivity; relations with donors and external collaborators; and levels of flexibility in grant management.

In some cases, it can be helpful to explore the extent to which collaborators and potential users are involved in implementing the research as they can often influence how the research evolves.

Stage 3: Primary outputs from research

The primary outputs from research fall into the first two payback categories: knowledge production and research targeting and capacity building.

Most knowledge production outputs (publications) will enter the pool of knowledge that informs further research, by the same or other research teams, either in the same field or in new research areas. They may eventually also become incorporated into policy-making or product development processes (see Stage 4, below). If the research helps to secure funding for subsequent projects, attracting new recruits or promoting researchers' careers, it will have contributed to building capacity.

Interface B: Dissemination

Dissemination is more than just the production of an academic publication containing results of the research. A crucial aspect of the translation process is the transfer of research findings to potential users in the political, industrial and professional environment and to wider society.

While citations of a published paper can serve as an indicator of dissemination and take-up within the academic world, numerous additional activities may also be employed to assist in disseminating research findings. These include conference papers and presentations, seminars, the production of audience-specific briefs, personal networks for knowledge exchange, educational activities, interactions with the media and so on.

Dissemination is a complex process that takes place over time – during, as well as after, completion of the research – and can reach beyond the immediate scientific sector to inform other research areas.

Stage 4: Secondary outputs – policy making and product development

Secondary outputs contribute mainly to the third payback category: informing policy and product development. Unlike primary outputs, secondary outputs can be difficult to identify and doing so requires the use of a variety of techniques, including interviews, database reviews and bibliometric analyses.

Evidence of any influence on policy can be found in the form of research citations in documents produced by professional and public policy-making bodies. Research findings can be used to develop new policy, change policy or maintain an existing policy at all levels within an organisation and with varying degrees of impact.

Secondary outputs can influence academic course curricula in medical schools and can be cited in patent applications or in the licensing of production rights for drugs, vaccines or medical equipment.

Stage 5: Adoption by practitioners and public

The adoption by practitioners of the outputs from research is important for their translation into widespread health and socioeconomic benefits. Adoption is generally accompanied by some sort of change in practitioner or public behaviour. Sometimes adoption comes as a direct result of primary outputs, for instance when clinicians decide to implement research findings before the development of clinical guidelines. Assessing the degree of public adoption of research is more complicated. One indicator is the extent to which patient behaviour changes as a result of interactions with those healthcare providers who promote research-based messages. Another might be the public response to media coverage of research findings.

When evaluating research outputs, it is important to try to establish adoption or take-up rates, and to explore how far a behavioural change can be attributed to the specific research findings, as opposed to other factors (such as a more general change in climate of opinion).

Stage 6: Final outcomes

The final outcome stage is reached when the broader health and economic benefits of research (the fourth and fifth payback categories) become apparent. These benefits can accrue over a protracted period of time and are the most difficult to attribute to an individual research grant.

2.5 Unit of analysis

One of the first decisions we had to make was which specific element of the projects making up the sample would be the focus of our study; in other words, what was our unit of analysis – the research grant, the scientist, the research programme or the host institution?

We wanted to identify, sample and follow individual pieces of research from inception through to the present day; but these needed to be limited enough in scope to allow us to construct detailed case studies around them. Our previous experience suggested that using research grants as the unit of analysis offered the best compromise. As well as being tractable in terms of size and relatively easy to obtain systematic information about, they are also of immediate relevance to funders – each grant is not only tied to a specific funder, but also to a particular funding decision. Using individual grants as the unit of analysis for the project therefore allows our analysis to reflect the choices that funders have to make and helps guide the relevance and applicability of the insights generated from the project. Grants also have the advantage of being easy to select because lists of grant awards are generally kept by the funding bodies that award them.

One disadvantage of using grants as our unit of analysis was the difficulty of separating the work supported by one grant from a group that has multiple, possibly overlapping, grants. This issue is discussed in more detail in the section on Case study scope on page 46.

2.6 Selection of time frame

In choosing an appropriate time window for Project Retrosight we had to balance the need to obtain reliable information with the need to allow enough time for the impacts of the research to develop. The longer the timescale we selected, the greater the danger that records would be missing and recall fading. On the other hand, examining recent research projects ran the risk that impacts would not have had a chance to develop.

Previous studies have suggested that principal investigators' (PIs') recollection of their research is generally very good, but that the extent of archive records available is heavily dependent on the practices of the funders (Wooding et al., 2004; Nason et al., 2008). Previously we had successfully carried out research looking back over a 10- to 15-year time period (Wooding et al., 2004). In consultation with the steering committee it was decided that we would attempt to look back over a 15- to 20-year period, so we selected a start point for grants between 1989 and 1993.

2.7 Selection of characteristics of interest and the selection matrix

Before embarking on our study we had to select factors that might potentially affect research impact and ensure that each was sufficiently represented in our sample. There are a great many factors that have been shown to influence research success, or may be assumed to do so, and in order to extract findings that could be generalised we needed to make sure that we had sufficient numbers to make a comparison along each selected dimension. Yin suggests that a useful heuristic for the number of necessary cases is four either side of a comparison (Yin, 2003).

We wanted our research to look not only at the specific factors that we had identified as being of particular interest, but also to identify and explore other factors that affect research success and describe how these might be influenced.

There are two types of selection criteria – those that represent dependent variables (measures of impact) and those that represent independent variables (things that might affect impact). Dependent variables included high and low impact; selecting both ensured that we had a variety of outcomes and so could explore what correlated with these. Independent variables, such as whether the research was basic biomedical or clinical, enabled us to focus on testing a particular hypothesis about whether this factor (which may be uncommon) has an effect on impact.

2.8 Factors used for selecting case studies

In consultation with Project Retrosight's steering group, we compiled a list of factors for comparative analysis. It included the following.

- **Country of research** (Australia/Canada/UK): selecting a balance of case studies across the three countries would allow us to look for national variations arising from differences in the research, health systems and cultures of the three countries.

- **Type of research** (basic biomedical/clinical): grants could be divided into two categories – those funding basic biomedical research and those funding clinical research. This would be based on the subject of the research grant rather than the qualifications of the researcher.
- **Impact** (higher/lower): research grants could be broadly separated into two groups – those that had achieved high impact and those whose impact had been low. Projects would be categorised based on data on the PIs' own perception of their impact, both academically (e.g. papers) and more widely, in terms of effects on health policy and health. This information would be collected through an initial PI survey.
- **Size of grant** (large/small): splitting the research grants according to the size of the award would permit us to look for differences between large and small grants.
- **Clinical experience**: dividing grants on the basis of whether or not their PIs had clinical experience or whether there was a researcher who had clinical experience in the team.
- **Academic age**: grants could be divided according to the number of years the PI had been active in research.
- **Age of independence**: grants could be divided according to the age at which the PI had started his or her own group, rather than carrying out research as part of a larger group.
- **Multidisciplinarity**: the range of disciplines in which the researcher, research group or grant was active could be used to categorise the case studies.
- **International collaboration**: the level of international collaboration on the grant could be used to categorise the case studies.
- **Soft/core funding**: the amount of soft, or discretionary, funding the grant PI controlled. There have been earlier indications that having such money allows researchers the freedom to pursue their instincts.
- **Research group culture or PI personality**: there is evidence that the personality of the PIs and the culture they develop in their research group can affect the wider impacts of their work.
- **Plurality of funding**: the number of different funding sources supporting PIs' research.

- **Critical mass:** dividing researchers by the size of their research groups – or possibly the institutions in which they are situated – would allow us to investigate the issue of whether scale, or critical mass, affects wider impacts of research.
- **Institutional setting** (university, institute, medical school or hospital): examining whether the location of the research group affects its impact.
- **Curiosity-driven versus mission-directed:** dividing research according to how much influence funders exerted on the direction of the research – from purely curiosity-driven through to directly commissioned research to answer a specific question.
- **Peer-review ranking:** dividing research according to how well it was rated in the peer-review system for funding application.
- **Type of research funding:** comparing project, programme, fellowship and/or institute funding.

However, the scale of the project (we had settled on a sample of 25–30 case studies) suggested that we could focus on only four specific factors for comparative analysis. From this long-list we isolated four feasible key factors that would potentially provide insights to funding bodies. We selected case studies that would balance the following factors.

- **Impact of research** – in order to get an immediate measure of the impact of each research grant, we aimed to select a sample that contained both high- and low-impact research. This would allow us to support our conclusions by demonstrating the presence of certain features in high-impact research and their absence in lower-impact research.
- **Type of research** – this allowed us to examine the relative levels and types of impact of basic biomedical and clinical research, something that was also a key interest of the steering committee.
- **Country of research** – Project Retrosight gave us an opportunity to examine whether or not national context plays a key role in research outcomes.
- **Size of research grant** – we wanted to try and understand how different sizes of grant were used and the relative benefits that each brought.⁴

To reduce the variety of other factors we needed to consider we decided to focus on just one mode of funding: project/activity grants. This was largely because we needed a funding type that was broadly similar, and therefore comparable, in all three countries. We have

⁴ In later analysis this proved problematic as the range of sizes was relatively small, differed between countries and was not always correctly reported in funders' records. For these reasons we have not attempted any analysis based on grant size.

3.1 **Introduction**

In this chapter we describe in more detail how we identified the grants that would form the subject of our case studies. Essentially, this process consisted of consulting the records of funders to identify relevant grants; establishing whether the PIs of these grants had published between 2004 and 2006; locating contact details for the PI; surveying the PIs to collect more information about them and their projects; and then using all this information to assign the grants to one of the cells of the selection matrix.

3.2 **Grant identification**

The selection process began with the identification of grants awarded in the cardiovascular and cerebrovascular research field between 1989 and 1993 by various funders, including the British Heart Foundation (BHF) and the Stroke Association in the UK, the Heart and Stroke Foundation (HSFC) and Canadian Institutes of Health Research (CIHR) in Canada and the National Heart Foundation (NHF) of Australia. All of these organisations were able to provide lists of grants awarded during the project time period. The National Institute for Health Research (NIHR) had no such centralised lists for its funding from the early 1990s because most of its funding was administered locally at that time, and so we were unable to select grants supported by NIHR. Had we needed to extend the range of grants from which to select our case studies, we might have consulted other organisations in the study countries that were funding research in the cardiovascular field during the relevant time period (these included the Medical Research Council and the Wellcome Trust in the UK and National Health and Medical Research Council in Australia). However, given the other sources of funding held by our case study PIs we are confident that our case studies cover the range of research in cardiovascular disease and stroke at the time, as well as a sample this size can.

We assumed that all grants from the BHF, Stroke Association, HSFC and NHF would be within the scope of the project. In addition we added grants from the CIHR that were considered to be in the field of cardiovascular and stroke research. If the CIHR research team considered the grant title to be relevant, a checking process aided by both CIHR's own funding data and information on projects co-funded by HSFC, the project was deemed to come within the scope of Project Retrosight.

3.3 PI survey

The funder records did not contain enough information for us to assign the grants we had identified to cells in the selection matrix (particularly for categorising by research type), so we decided to collect the extra information by carrying out an online survey of the PIs.

Broadly speaking, the survey content was the same for all three countries, although we made some minor adjustments to the language and terminology to take account of particular characteristics of the research system in each. We also provided a French translation of the survey in Canada.

3.3.1 Survey outline

At this early stage we were relying on the willingness of the PIs to get involved with the project. To maximise response rate and avoid burdening PIs, it was crucial that we kept the survey concise.

The survey had three sections: the first asked for basic information about the PIs, including their gender and the number of years, post-doc, they had been working in research. The second and third sections asked about grants awarded to the PI in the period 1989–93. If the PI had received more than one grant, the second and third sections were repeated for the first two grants awarded. We did not ask about more than two grants to minimise the time needed to complete the survey and (we hoped) increase the response rate.

The second section focused on the nature of the grant itself, including whether it was concerned with basic biomedical or clinical research and whether it was interdisciplinary in focus. We had initially hoped to be able to separate the basic biomedical research grants from the clinical ones using only funders' records; unfortunately, we found that each of our funders had a slightly different way of distinguishing between the two. Because of this we asked the PIs to classify their research according to the following hybrid of our funders' definitions:

- basic biomedical – focuses on normal or abnormal function at the molecular, cellular, organ or whole body level;
- clinical – focuses on patients, better diagnostics or treatments and increasing quality of life.

The third section of the survey asked about PIs' perceptions of the impact of their research projects. This included information about the number of publications directly attributable to the grant, whether or not the research had led to further work by them or others, whether any post-graduate degrees were attributable to the grant, the scale of the project's impact on policy, whether or not the research had led to product development and an assessment of the project's impact on clinical practice.

The survey was designed as a simple tick-list that would take no more than ten minutes to complete. Most questions required a "yes/no" answer while questions such as the extent of impact on clinical practice were answered using a three-point scale (Appendix A gives the full survey). These techniques made it easy for PIs to complete the survey and easy for us to add up the overall perceived impact of the grant.

3.3.2 Building the survey tool

As our survey sample was spread over three countries, we felt the most efficient way of conducting the survey would be online. Because of the highly specific nature of the data collection, and the fact that only a small number of coded answers were required, we built our own online tool to collect the survey responses. This had the added benefit of further reducing the administrative effort required to complete the survey by the PIs, since each individual could be sent a unique URL.

3.3.3 Testing the survey tool

Before using the survey, we conducted preliminary tests with PIs in each of the participating countries and requested feedback. The response was generally favourable, but at the same time some significant criticisms were made that enabled us to fine-tune the content and structure of the survey.

Some of the test respondents believed that the survey was the main method of data collection for the whole project. We responded by inserting a line in the survey introduction to explain why the information was being collected and that the focus of the research would be detailed case studies of the grants in question.

It later transpired that our test-run failed to test the data extraction from the survey sufficiently, with the result that some grants ended up wrongly classified in the initial selection matrix. We were able to correct this mistake subsequently but it did mean that some case studies that had already been started had to be reclassified. This contributed to some slight imbalances in the final case study matrix.

3.3.4 Survey deployment and response rates

To carry out the survey we had to obtain current email addresses for all the PIs. For Canada and Australia the funders were able to provide us with relatively complete, up-to-date email lists (covering 80% and 99% of PIs in our time window respectively). However, in the UK such lists were not available and we had to hand-search for email addresses, identifying them for 47% of UK PIs. To focus our search effort, we excluded PIs who had not published between 2004 and 2006. To identify PIs’ publications we searched the Thompson Reuters *Web of Science* databases using the researcher’s surname and first initial. We used only one initial, which increases the chances of including some papers in fact not by the PIs, because we wanted to err on the side of inclusion rather than exclusion. This filter excluded 21% of the PIs initially identified. Table 3.1 gives more details on the number of email addresses and bibliometrically active PIs.

Table 3-1: The number of PIs, email addresses and level of bibliometric activity

	Australia	Canada	UK
Number of PIs identified from funder lists	222	572	553
Number of PIs for whom we had email address	220	457	261
Number of bibliometrically active PIs – published in 2004–6	180	427	451
Percentage of identified PIs bibliometrically active	81%	75%	82%

3.3.5 Identifying email addresses

We located email addresses for the bibliometrically active researchers from funder records and through the PubMed online database and Google searches. We then surveyed these PIs to find more details about them and about the grant. Some PIs had been awarded several grants during the time period specified, but we asked them to provide answers for no more than two in order to keep the survey concise. The average response rate to the survey, in terms of grants, was 33% and was similar across countries (see Table 3.2). This raises the possibility that some difference between responders and non-responders may have biased our sample. In particular, it might be expected that those researchers who perceived their projects as having a greater impact would be more likely to respond. Our deliberate selection of both high- and low-impact research should mitigate this possibility to a large extent, but it is nonetheless a caveat to bear in mind when considering the place of our case study grants in the wider research context.

Table 3-2: The number of PIs contacted and the number who responded in each country

	Australia	Canada	UK
Number of PIs surveyed	220	457	261
Number of responses	72	158	78
Response rate	33%	35%	30%

4.1 **Filling the matrix**

Having decided on the time window, selected our key characteristics, identified relevant cardiovascular and stroke research grants awarded during that period and collected the necessary additional information on each grant through the PI survey, we then had to make our final selection of grants to investigate in case studies.

To do this, we allocated each grant to a cell of the selection matrix thus.

- **Country:** allocated according to the country in which the research was carried out, even if the researcher had subsequently moved countries.
- **Type of research:** allocated to clinical or basic biomedical according to the PI's survey response.
- **Impact of research:** calculated from the PI's survey response by simply totalling the responses in the impact section of the questionnaire. We plotted the distribution of impact across all grants (separately for basic biomedical and clinical research) and then cut these distributions in half to produce higher impact (top half) and lower impact (bottom half).
- **Size of research grant:** from funder records. We removed the extreme outliers and then split the distribution in two, considering the top half "larger grants" and the bottom half "smaller grants".

From the data collected in the survey we were able to populate our case study selection matrix (see Table 4.1)

Table 4-1: Number of grants we received survey responses on in each cell of the selection matrix

	Canada		Australia		UK	
Basic biomedical	Higher impact	71	Higher impact	27	Higher impact	21
	Lower impact	60	Lower impact	28	Lower impact	24
Clinical	Higher impact	7	Higher impact	5	Higher impact	7
	Lower impact	14	Lower impact	5	Lower impact	8

4.2 **Selecting the cases**

Having generated a list of grants for each cell in the matrix we then worked through the cells, country by country, randomly selecting (by coin-tossing) which grant to use for the case study.

Analysis of our survey results showed that 25% of the responding researchers were women. We therefore tried to select so that 25% of the case studies had female PIs to ensure there was no risk of gender bias. We did this by working from country to country and, if necessary, selecting randomly from just the women or men in a particular cell.

In Australia and Canada we were asked to select a geographical distribution of case studies. We did this by monitoring the emerging selection and, when necessary, selecting from only the case studies in under-represented geographical regions from particular cells.

We then approached PIs to participate in the study. In instances where PIs declined to participate in the research or were otherwise excluded, we returned to the cell in the matrix and selected randomly from the remaining case grants. PIs were excluded from the research if they had moved from the initial country of the research to a country outside the study (this happened in one instance). There were no examples of PIs moving from one of the countries participating in the study to another after their case study grant, although some had moved between participating countries before the period surveyed.

We had agreed to exclude PIs who were in legal dispute with any of the funders, but there were no instances of this in our selection.

5.1 **Pilot case study**

Before embarking on the study we conducted a pilot case study to test our methodology. In a previous study for the Arthritis Research Campaign (Wooding et al., 2004) we undertook an initial round of case study pilots and reported these to our steering committee before proceeding with the subsequent 13 case studies. These initial studies were valuable in demonstrating the applicability of the Payback Framework to early clinical and basic biomedical research and also provided methodological refinements, for example highlighting the importance of interviewing the PI early in the process, before extensive review of subject-specific documentation.

Although our previous work gave us confidence that the same approach would work for Project Retrosight, we wanted to confirm our belief that the Framework was applicable to cardiovascular and stroke research. Therefore we conducted a single pilot case study in the UK and used the results, along with an explanatory research pack, to illustrate the approach to the study teams in Canada and Australia.

While we did not consider it necessary for the other participant countries (Canada and Australia) to conduct additional pilots, we staggered the start of the case studies to allow early feedback and to give ourselves time to make any necessary adjustments. The case studies were therefore separated into two tranches that began successively rather than simultaneously.

Selection of the pilot case study was largely pragmatic because we wanted to study a researcher whose work was likely to have had some impact and who was easy for us to visit. We began the pilot exercise by approaching specialists in the field of cardiovascular medicine to get an idea of the current trends in cardiovascular and stroke research. We then sought a suitable candidate to approach for our pilot study.

We selected a project run by a senior cardiologist based in Cambridge. The PI was an influential figure in the field of cardiovascular medicine, who had a reputation for producing high-impact research.

The case was a British Heart Foundation fellowship dating from 1990 and the PI was a practising cardiologist who had been encouraged to apply for the fellowship by a BHF researcher.

Having identified the pilot case we applied the framework and looked at the project inputs (money, knowledge, expertise, space, consumables and so on) and its outputs, including dissemination and final outcomes.

This pilot case study not only demonstrated the suitability of our chosen methodology but also provided us with a better understanding of how we should perform the actual case studies for Project Retrosight. It focused our attention on issues such as how to attribute research outputs to a grant, how to define an outcome, what constitutes adoption and what should be included as a “wider” outcome.

5.2 **Cross-country working**

The case study work for Project Retrosight combined several distinct research methods (described on page 25) and was carried out by teams of researchers in each country, coordinated by the UK team.

The Australian research was conducted by an internal staff member of the NHF of Australia and by market research specialists Instinct and Reason. In Canada, the research was carried out internally by staff at the CIHR. The UK team from RAND Europe and HERG conducted all UK research as well as coordinating the research carried out in Australia and Canada.

We took a number of steps to develop a consistent approach across the teams, including providing written protocols and templates and running a series of three workshops during the course of the project. The protocols are included as appendices to this report and the content of the workshops is described below. The workshops served as opportunities for the Canadian and Australian researchers to familiarise themselves with the Payback Framework and methods used in the study, and to review initial findings with the RAND/HERG researchers.

Workshop 1: Santa Monica, February 2007

- Familiarisation with different countries’ funding systems
- Introduction of the Payback Framework; presentation of pilot case study
- Discussion of process and data requirements for case study selection
- Discussion of project outline

Workshop 2: Santa Monica, December 2007

- Review of selection survey responses
- Review and discussion of initial case study work
- Review of case study templates
- Discussion of classification of narrative, outputs and outcomes into the Payback Framework

Workshop 3: Cambridge, June 2008

- Review of later stages of research for initial tranche of case studies and initial research on further case studies
- Further discussion of classification of narrative, outputs and outcomes into the Payback Framework
- Review of challenges and issues arising from data collection and case study write-ups
- Presentation of initial bibliometric analysis for case studies
- Discussion of structure of analysis phase

5.3 Case study research methods

Each case study combined data from a variety of sources including archival documents and published literature, bibliometric information and semi-structured interviews with PIs and other informants. In this section we look in detail at how the case study research was carried out.

A range of background material was reviewed before conducting interviews for the grants selected, and this was referred to throughout the case study research process.

5.3.1 Document and literature review

We undertook a review of a range of materials – original grant applications, referees' reports and other relevant documentation (including subsequent correspondence between applicants and funding bodies) – to gather information about the grant and the PI. We also reviewed interim and end-of-grant reports if they were available.

In addition, we skimmed the peer-reviewed papers and reports that had been produced by the PI at around the time of the grant. These documents were located through searches in Thompson Reuters *Web of Science* databases and Google Scholar. After we had carried out the PI interview we then reviewed in detail papers the PI considered to have arisen from the grant.

The pilot case study demonstrated that even if the funder had disposed of grant records, the PIs or their institutions may have retained copies of such documents.

5.3.2 Interviews

After our background research we conducted semi-structured interviews with the PI for each grant. Through these interviews we identified others who had also been involved in the research (clinicians, collaborators and other stakeholders), and they were approached for interview as well. Conducting multiple interviews enabled us to obtain several perspectives on the issues we were investigating and gave us enhanced insight into individual experiences. Different interview protocols were employed for PI and non-PI interviews (see Appendices B and C).

The interviews were modelled on the Payback Framework and explored the origins of the research, the primary outputs and any translation of research findings into product development, policy or practice.

Most interviews were conducted face-to-face, with a few by telephone. Each interview lasted between 60 and 150 minutes. In addition, interviews were often followed up with email correspondence and telephone enquiries and clarifications. Interviews were recorded and the recordings preserved for later reference. All recordings were held in accordance with the retention and data protection policies of the organisation leading the case study.

The interviews began with an introduction to the payback model in the evaluation framework and an explanation of the aims of the study. The interview protocol followed the structure of the model itself. We explained to the interviewees that not all the questions would be relevant to their research project. The interviewers, during their desk research, had the opportunity to introduce additional questions and these were added to the basic interview protocol.

Non-PI interviews also closely followed the payback model but differed from the PI interviews in that they reflected the informant's involvement in the project. Non-PI interviews were often more focused than PI interviews, concentrating on a specific area of the research or dissemination. For example, if informants were only involved in dissemination, they were not asked about the early stages of the project.

5.3.3 Bibliometric analysis

Bibliometrics is the use of publication and citation data to assess the research performance of individuals, groups and institutions. We used this method to assess the significance of the papers produced from our case studies. To do this we identified the papers directly and indirectly linked to the case study grant and analysed the citations that they had received. This required us to carry out two key steps: first, identifying the relevant papers and second, capturing the citations of those papers.

On occasion we also used bibliometrics to trace the impact of case study research by approaching authors who had cited the papers to ask how it had affected their research.

Identifying the papers

Initially we identified lists of publications for all PIs from the year in which their research grant started to the present day. Details were extracted either from the PI's CV or from the Thompson Reuters *Web of Science* (WoS), which includes author lists, article titles, publication names, volume and issue numbers, page numbers and finally ISI (Institute for Science Information) numbers in each case.

We showed this list to PIs during the interviews and asked them which papers could be attributed to the case study grant. We asked them to classify related papers into two groups:

- directly related – papers that arose directly from the research conducted on the grant in question;

- indirectly related – papers to which the research on the grant in question made a substantial contribution, and to which other research grants held by the PI may also have contributed.

We also asked the PIs to add other publications (such as abstracts and poster presentations) that were not in the WoS database or noted on their CV.

In some cases there was confusion over definitions. For example, in a few instances the number of papers the PI claimed were related to the grant was so large that it raised doubts about the legitimacy of the claim. In others, the PI and non-PI interviewees disagreed about the attribution of particular papers.

For instances where there was confusion over the attribution of papers, we developed a set of guidelines to help the Project Retrosight case study research teams identify whether papers should be regarded as attributable, directly or indirectly, to the grant in question (see Figure 5.1).

Figure 5-1: Guidelines for paper attribution

We provided the following guidelines to help case study researchers decide whether or not papers were attributable to the grant in question:

1. Does the paper acknowledge the specific grant (e.g. through a grant number)?
2. Does the paper acknowledge funding from a body other than the grant funder?
3. Is there common authorship between the grant applicant (PI, co-PI or other named investigator) and the paper?
4. Is there common keyword or title use between the grant application and the paper?
5. Is there common content between the grant application and the paper?

If the answer to question 1 was “yes”, the paper was attributable to the grant.

If the answer to question 2 was “yes”, the paper was not attributable.

If the answers to questions 3, 4 and 5 were “yes”, the paper was attributable

Final paper lists were reviewed by the country teams to ensure they were accurate and the list of papers classified as directly attributable was analysed. Although some discrepancies remained, it is important to view these in context: the bibliometric analysis provided only

one strand of evidence for the level of knowledge production, which was itself only one strand of data for the final analysis.

Analysing the publications list

The extraction of bibliometric data used in Project Retrosight was undertaken by the Centre for Science and Technology Studies (CWTS) and the data were analysed by the Research Evaluation and Policy Project (REPP) at the Australian National University.

The bibliometric data detailed in Figure 5.2 were extracted by CWTS to assist in assessing the knowledge production of each of the case study grants. The data were presented in the form of a one-page summary for each case study. Once the paper lists had been compiled and agreed, the papers were identified in the CWTS database and their citations analysed. Details of the analysis methods are given in Figure 5.2, followed by descriptions of the indicators computed.

Figure 5-2: Details of analysis methods and indicators used

<i>Analysis methods</i>	
Database used	
The Centre for Science and Technology Studies (CWTS) at the University of Leiden bibliometric database, which is based on the Thomson Reuters <i>Web of Science</i> (WoS) database.	
Period covered	
1981–2007 for the overview analysis; year of publication to 2007 for grant-related publications.	
Counting method	
Whole counting, i.e. a full publication and citation count was allocated to grant outputs even when the publication was co-authored with researchers outside the study group.	
Publications covered	
The analysis included all research and review articles, notes and proceedings papers in journals indexed in WoS. Letters to the editor, editorials, and the like were not included, and articles in journals not indexed by WoS were also excluded from citation analyses.	
<i>Indicators</i>	
Number of journal articles	
Total number of journal articles directly related to the case study research project.	
Number of articles indexed by WoS	
The number of these articles appearing in journals indexed by Thomson Reuters in the WoS database.	
Total number of citations (all papers)	
The number of citations received by these (WoS-indexed) articles from the year of publication to the end of 2007.	
Aggregate relative citation impact	
The average citation per publication rate for the grant outputs was compared to the citation per publication rate for all cardiovascular/stroke publications, calculating a relative citation impact (RCI). This overall RCI was then sorted into one of five classes:	
Class I	uncited
Class II	> 0 but < 0.8
Class III	0.8 to 1.2
Class IV	1.2 to 2.0
Class V	> 2.0
Self-citations	
Self-citation was defined as the proportion of citations received by the grant-related publications that came from other publications written by one or more of the same authors. Although self-citations were included in citation counts, the extent of their contribution to the total citation count was calculated and indicated in the case studies.	

Distribution of publications across relative citation impact classes

Each publication derived from the grant was assessed against the cardiovascular/stroke citation per publication rate for the relevant year, and assigned an RCI class. The publications were then distributed across the five classes in tabular form.

The most highly cited publication in each case study was identified, and final citation counts were extracted for these in April 2009.

Fields of research of output

The publications related to each grant were classified, on the basis of the journals in which they appeared, in one or more of the WoS subject categories. The data were then presented in the form of a pie chart.

Fields of citing papers

All articles that cite the publications related to a grant were classified, on the basis of the journals in which they appeared, in one or more of the WoS subject categories. In the WoS classification scheme, some journals are classified in more than one field. Where this occurred, the count for the article was fractionated across the fields. The data were then presented in the form of a pie chart.

Countries of citing papers

Data on the countries of the citing articles (determined from the addresses attached to the articles) were collated. Where authors from two or more countries collaborated on one of the citing papers, the count for the article was fractionated across the fields.

5.4 Case study write-up and review

The different types of information collected for each case study were written up using a standard template structured around the Payback Framework. The three international workshops, held during the write-up of the first tranche of case studies, helped to standardise our reporting and allowed us to refine the templates.

To improve consistency between countries further, two members of the RAND/HERG team reviewed all the initial drafts of the case studies and provided feedback on tone, areas for further examination, and inclusion and exclusion criteria for impacts and outcomes. The case studies were then completed by the case study research teams. Once each case study had been drafted it was circulated to the PI for clearance.

5.5 External review

The case studies depended heavily on information from the PIs and other informants involved in the research projects, much of which could be checked against documentary sources. But in order to ensure reliability we also sought independent external peer review. For each case study we identified other experts in the subject area and asked them to review the case studies for historical accuracy. These reviews were used in assuring the quality of the case studies and informing the subsequent impact rating process (described below).

Our ideal was to recruit two reviewers for each case study, one from the same country as the PI and one from one of the other two countries taking part in the study. We managed to achieve this for 24 of the 29 case studies but, despite approaching over 200 reviewers in total, we managed to secure only one reviewer for each of the five remaining case studies (one in Australia, one in Canada, three in the UK).

We identified potential reviewers in a number of ways: journal editorial boards, scientific advisory boards of medical charities and research funders, authors of highly cited reviews in the relevant field, and senior researchers in high-profile research institutes or university departments. We focused on senior people, because of their expertise and respected position, and also because junior researchers were unlikely to have been actively involved in the field during the case study period. If they agreed to participate, we sent reviewers the draft case study and asked them to comment on the accuracy of both the impacts attributed and the science described, as well as whether the influence of others in the field had been sufficiently acknowledged. The instructions and review form provided to reviewers are included in Appendix D.

The verdict of the majority of reviewers was that the case studies were a reasonable summary of the scientific history and the role played by the case study PI. In only one instance did we receive dramatically conflicting reviews; for this case study we sought a third opinion to clarify the situation. Summary tables and the full text of the reviews were used during the analysis, described in the following section.

The questionnaire we circulated to PIs provided an initial subjective measure of the impact of their research; but after completion of the case studies we had far more information, from a variety of sources, on which to base this assessment. In this section we describe in detail the process we used to rate the case studies, including how we assembled and deployed the rating panel, through to the rating scale we used and how we carried out the rating process.

Inevitably, each grant had achieved different levels of impact in each of the five categories of the payback model. So as well as assessing the overall impact and value for money of each grant, we also wanted to assess the level of impact for each in each individual payback category.

Each member of our rating panel was asked to produce a series of ratings for each case study, covering the five payback categories (knowledge production, research targeting and capacity building, informing policy and product development, health and health sector benefits, broader economic benefits) plus overall impact and value for money.

6.1 **The rating panel**

The panel was made up of both researchers familiar with the project and others new to it, who were invited to provide a degree of external validation. It included a case study researcher and an external expert on biomedical research funding from each project country, as well as one expert from a country not included in the project to provide an entirely independent perspective. The full membership of the panel is shown in Table 6.1.

Table 6-1: Rating panel membership

	Expertise
Jonathan Grant	Study co-principal investigator and methodology expert
Martin Buxton	Study co-principal investigator and methodology expert
Stephen Hanney	UK case study researcher and methodology expert
Rob Mercer	Australian case study researcher
Laura McAuley	Canadian case study researcher and expert on Canadian biomedical research funding system
Liz Allen	Methodology expert and expert on UK biomedical research funding system

Marcus Nichol	Expert on Australian biomedical research funding system
Cy Frank	Expert on Canadian biomedical research funding system and methodology expert
Brendan Curran	Expert on Irish biomedical research funding system and methodology expert

6.2 Rating scale

In order to reduce each case study to a set of ratings for each of the five payback categories, for overall impact, and for value-for-money, we devised a rating system based on a scale from 0–9:

0 = no impact in this category

9 = the case study (or studies) in the complete set with the most impact in the category

1 = the case study (or studies) in the complete set with the least impact in the category

The remaining case studies were given ratings between 0 and 9 based on their relative impact in the category. Multiple case studies could receive the same rating (so, for example, more than one case study could be rated 4 in any given category). Similarly, there could also be gaps in the 0–9 scale (there may be no case study rating 6, for example).

This scale was derived from the RAND/UCLA appropriateness method (Fitch et al., 2001), previously used for our work on arthritis research, with the addition of 0 for no impact.

The system had a number of advantages. First, the scale was both discriminating and manageable: the 0–9 scale provides a reasonably sensitive measure compared to, say, a 0–3 scale, but is easier to apply than a ranking of all 29 case studies.

Second, the rating system ensured the use of the entire 0–9 scale. Raters were required to use the whole scale, rating at least one case study 0 or 1 (for no impact or least impact) and at least one 9 (for most impact). Had we not done this, the ratings could have been distorted by subjective judgements; for example, a rater might consider that none of the cases was worth a 9 so the range of ratings used would be narrowed.

By reserving the value of 0 for no impact we ensured that all the case studies with impact were spread over the entire range of the scale (1–9), rather than allowing each rater to decide what the case study with the least impact above 0 was worth (in such instances, one person's 2 may be another's 5). This did have a disadvantage, in that it meant that the range of the rating scales in different categories were not comparable. For example, a 9 in one category did not necessarily represent as great an impact as the same rating in another category. However, judgements regarding the relative value of impacts in different categories, which are by definition qualitatively different, would be a highly subjective exercise, and we felt that there was greater value in obtaining a more fine-grained analysis within each category by spreading the ratings out on each scale as much as possible.

Our rating system also facilitated the presentation of agreement and disagreement among the raters. The scale allowed us to draw a simple histogram to show how well ratings agreed and to quantify this relatively simply by measuring the mean deviation from the median – that is, the average of how far each person’s rating was from the median rating (the calculation of this is discussed further in section 6.5). Such a process would also allow us, if necessary, to moderate the ratings of individual raters. Comparing the behaviour of individual members of the rating panel, by plotting the distributions of their ratings (overall or by category), allowed us to see if some were consistently more conservative or more generous than others.

The rating scale also ensured independence of ratings. Members of the rating panel could change the rating of one case study in a category without affecting their ratings for other case studies. Had we used a system of rankings, moving one case from 2 to 8 would have affected the rankings of all the case studies in between.

6.3 **Rating material**

The panel was provided with summary material prepared from the case studies but they also had access to the full case studies and peer review comments in case they wished to refer to them for more detailed information.

The set of summary material included the following.

Payback table

This summarised the key outputs for each case study under each of the five payback categories. Raters reported that this table was the primary tool used in the rating process.

A case study summary table

This table summarised the case study narratives organised by stage of the research process. Each section could be cross-referenced back to the full case study text and, as well as providing additional background material, gave a fuller explanation of the context than the payback table.

Bibliometrics data

These were presented as a series of charts showing the number of papers attributed to the grant, the number of citations they had received and how these citations were distributed.

Impact matrix

This was a graphical representation of the impacts of the grant covering the two payback categories research targeting & capacity building and informing policy & product development, presented as a matrix. The representation was a simplification of the RAISS impact array representation and was completed by the case study authors (see Wooding et al., 2009 for further details on the RAISS impact array).

Peer review summary table

This table summarised the comments received from peer reviewers on the accuracy and clarity of each case study, with particular focus on comments about the level of impact and degree of leadership in the field.

6.4 Rating process

The panel was then asked to complete the rating process over a two week period in advance of the February 2010 rating workshop, where the process – and some specific rating decisions – were discussed. The main aims of bringing the rating panel together were to ensure we had a common understanding of the rating process and to give us the opportunity to re-rate some of the case studies in respect of any improvements in understanding that came about through discussion at the workshop. No efforts were made to reach consensus scores; the re-rating simply allowed raters to consider information heard at the workshop and integrate it into their ratings.

The aim of the rating process was to define groups of high-, mid- and low-impact case studies in which we could identify and test factors that may be associated with impact. As such, the precise scores themselves were not critical; what was important was the position of each case study relative to the others within each category. High- and low-impact groups could then be compared to a series of binary variables representing the presence or absence of each factor in each case study. Grouping case studies in this way and focusing on the two extreme subgroups meant that we did not use all the data at the initial stage of analysis. However, this is a well-established technique in scale development, and division into three groups to conduct such analysis is in line with the standard method for computing classical discrimination indices (e.g. Crocker & Algina, 1986). Computing point-biserial correlations, or non-parametric equivalents, on the entire data set would have been less feasible given the small number of cases involved, and may have given the impression that our rating figures had an unwarranted level of accuracy. Further detail on how the impact categories and groups were derived is provided in Chapter 8.

6.5 Initial rating

Each panel member was provided with an instruction sheet (see Appendix E), the set of summary material and a scoring sheet, and asked to rate all 29 case studies in each of the seven categories. All members completed their rating according to the instructions and prior to the workshop.

In order to assess the rating process and prioritise which case studies to discuss at the rating workshop, we measured the degree of disagreement between raters. This was done by taking each payback category in turn and, for each case study, calculating the deviation from the median for each rating and averaging them. We termed this measure the Average Deviation from the Median (ADM). It is similar to the inter-rater reliability measure Cronbach's Alpha, which is based on the mean, but we chose to use the median because it is less sensitive to extreme values and shape of distribution. In addition to its neutrality to the latter, it is also intuitively appealing and easy to understand from a rater's point of view. It is calculated as follows:

$$Index = \frac{\sum |score - Median|}{N}$$

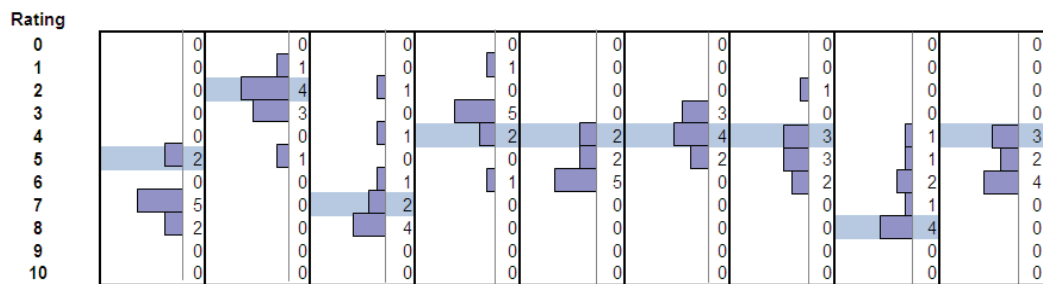
We used these case study/category ADMs to help us decide which case studies to discuss in each category at the rating workshop.

6.6 Workshop

At the workshop we worked through the ratings category by category, discussing the cases in which there was most disagreement. The aim of the workshop was not to alter the opinions of any of the panel members or to produce consensus, but to reduce differences in understanding of the projects or impact terminology while preserving genuine differences in subjective valuation of the impacts.

Each member of the rating panel was given histograms showing where their rating fell in the overall distribution of ratings but not revealing who had produced any of the other ratings, as shown in Figure 6.1.

Figure 6-1: Illustrative section of rating sheet



This system was designed so that no individual member of the panel was singled out as having rated differently from the rest. We did this to avoid forcing a consensus because we wanted to preserve differences in valuation.

Time constraints meant we could not discuss every rating so we concentrated on those ratings where there was the highest level of disagreement or where raters had expressed a particular desire to discuss a certain case study.

CHAPTER 7 **Data characterisation and robustness testing**

Before moving on to our detailed analysis we wanted to understand the data we had produced and explore potential biases in our analysis.

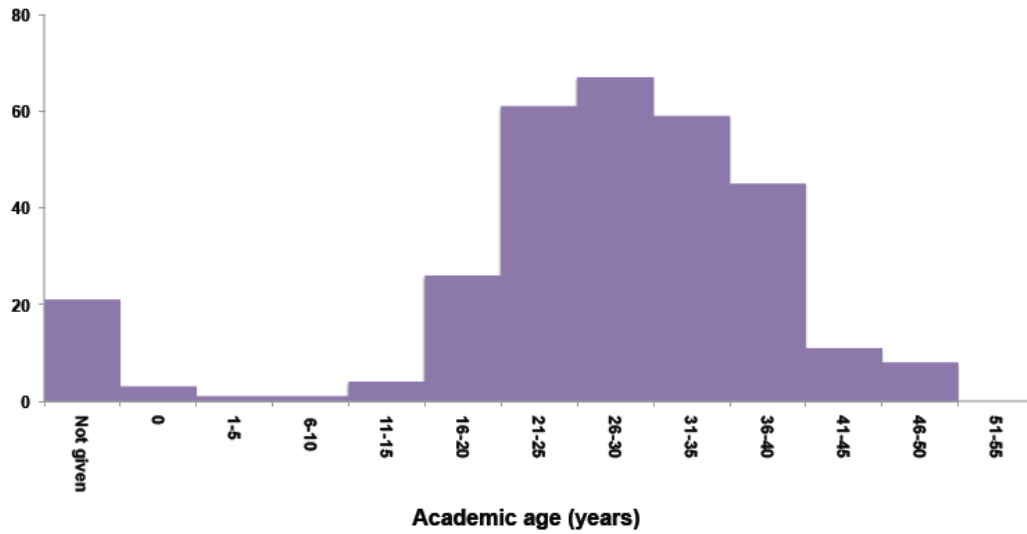
In any project like this there will be biases and weaknesses in the data set. In this section we review the tests we did to reassure ourselves about the robustness of our data, examine the features that might have confounded our observations and explain how we dealt with them. We examined various features of our data, starting with the characteristics of the case study sample, and then looked at the rating process to check whether our panel members were behaving similarly. Finally, we examined characteristics of the case studies that we thought might confound our analysis, looking at the scope, basic biomedical/clinical definition, and presence of negative or null findings.

7.1 **Potential sample bias**

As discussed previously in Chapter 3, there may have been some unavoidable bias in the response to our PI survey. In addition to this, we identified two other ways in which our selection methodology could have created bias in our final case study sample.

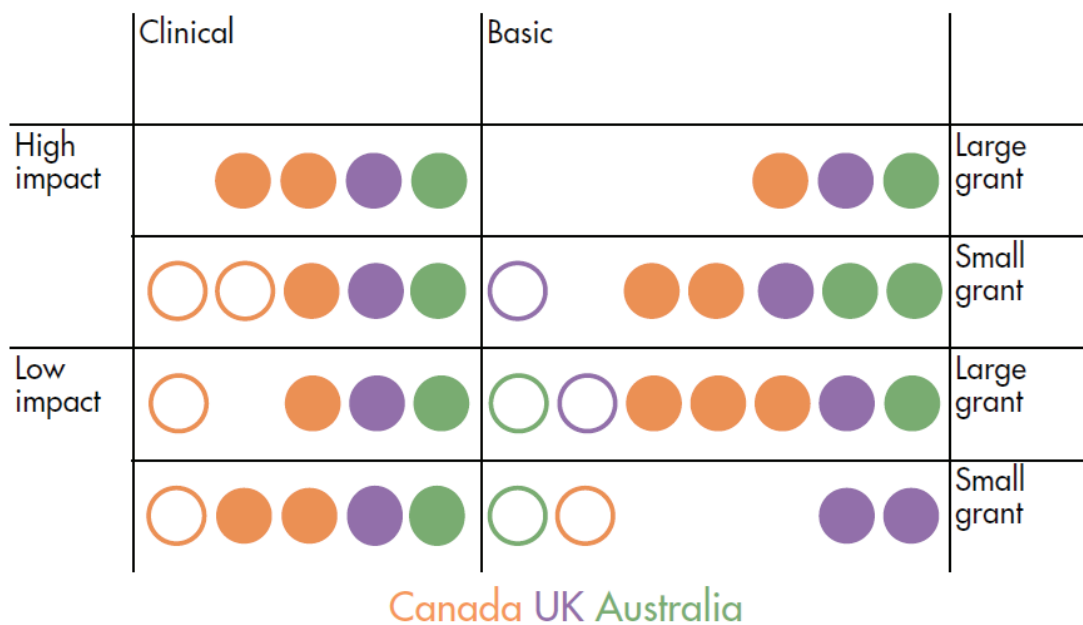
First, in relying on a bibliometric filter to focus our search for PIs, we are likely to have biased our sample in favour of younger PIs and those who remained in research. This could have led to more successful grants being overrepresented in our sample, as researchers who are successful are more likely to have stayed in research. Clearly we cannot eliminate this bias; however, examining the academic age distribution of the sample, shown in Figure 7.1, the number of years that PIs in our sample had spent in research was distributed across a wide range. Further reassurance about the diversity of our sample is provided by the variety of stages our PIs had reached in their careers at the time the case study grants commenced: these ranged from pre-PhD through post-doc to lab head and head of department.

Figure 7-1: Total number of years the PIs responding to our survey had spent in research by 2007



The second point at which we risked biasing our sample was through those PIs who were excluded from or declined to take part in the study. The distribution of these exclusions across the selection matrix is shown in Figure 7.2. Six PIs declined to participate in the study; in one other case initial scoping following the PI’s agreement to take part revealed that the grant did not meet our inclusion criteria. In one instance the case study research was initiated, and although three interviews were completed, insufficient evidence was available to allow full compilation of a full case study report. This case was thus excluded from the sample. Finally, in one case the PI had moved to Japan and hence a face-to-face interview was not feasible.

Figure 7-2: Distribution of exclusions: completed case studies are shown as filled circles, refusals and exclusions as open circles

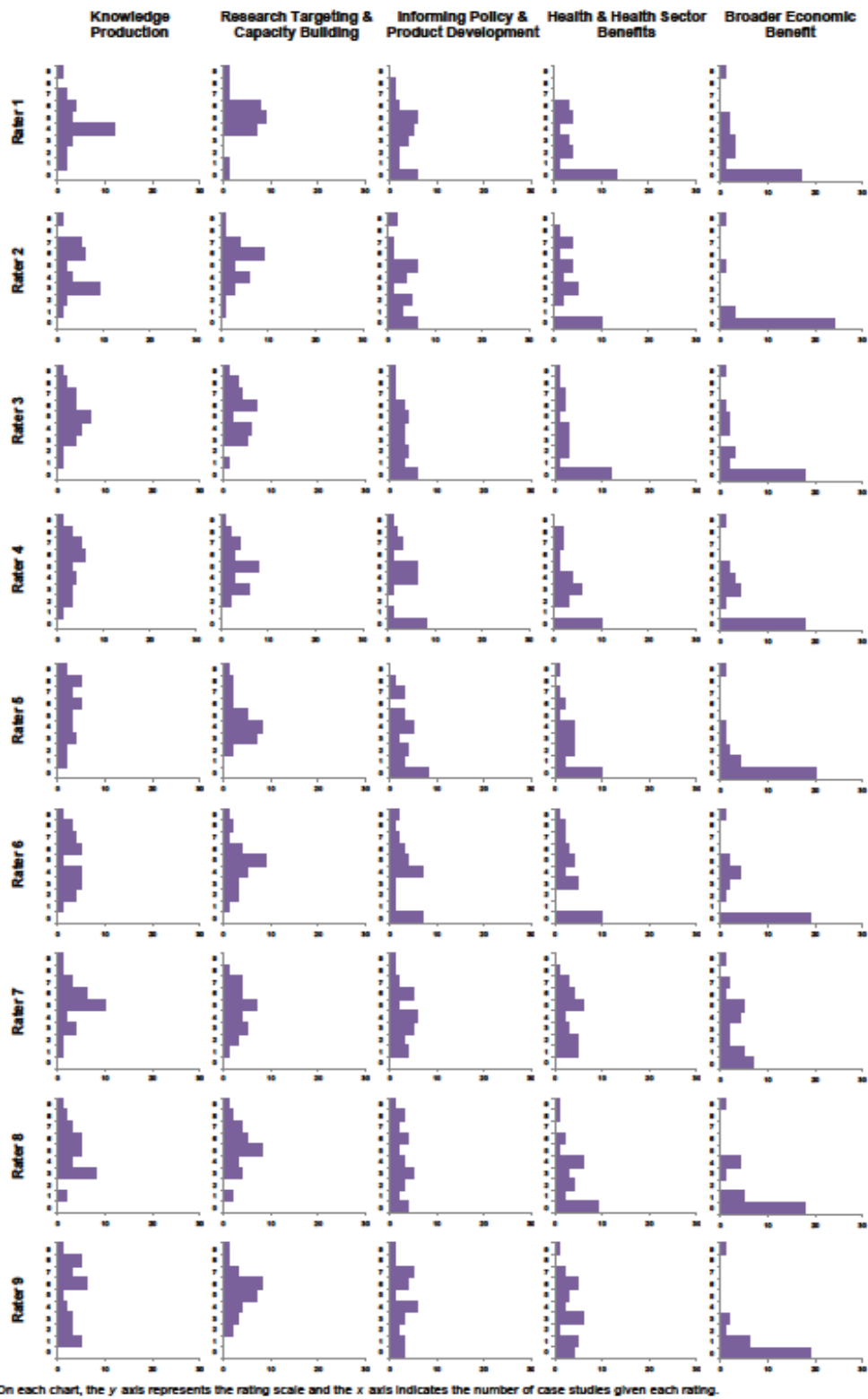


7.2 **Rating behaviour**

Because the ratings produced by the rating panel were a key source of data, it was important to examine how much agreement there was among panel members and whether they showed similar rating behaviour – i.e. whether some were consistently more generous than others.

First we examined the raters' behaviour, using the 1827 ratings from the first round of rating. It is clear from examining the distribution of the ratings that all the panel members behaved similarly in terms of the distribution of their scores. This is shown in Figure 7.3.

Figure 7-3: Ratings distributions (by person and category)



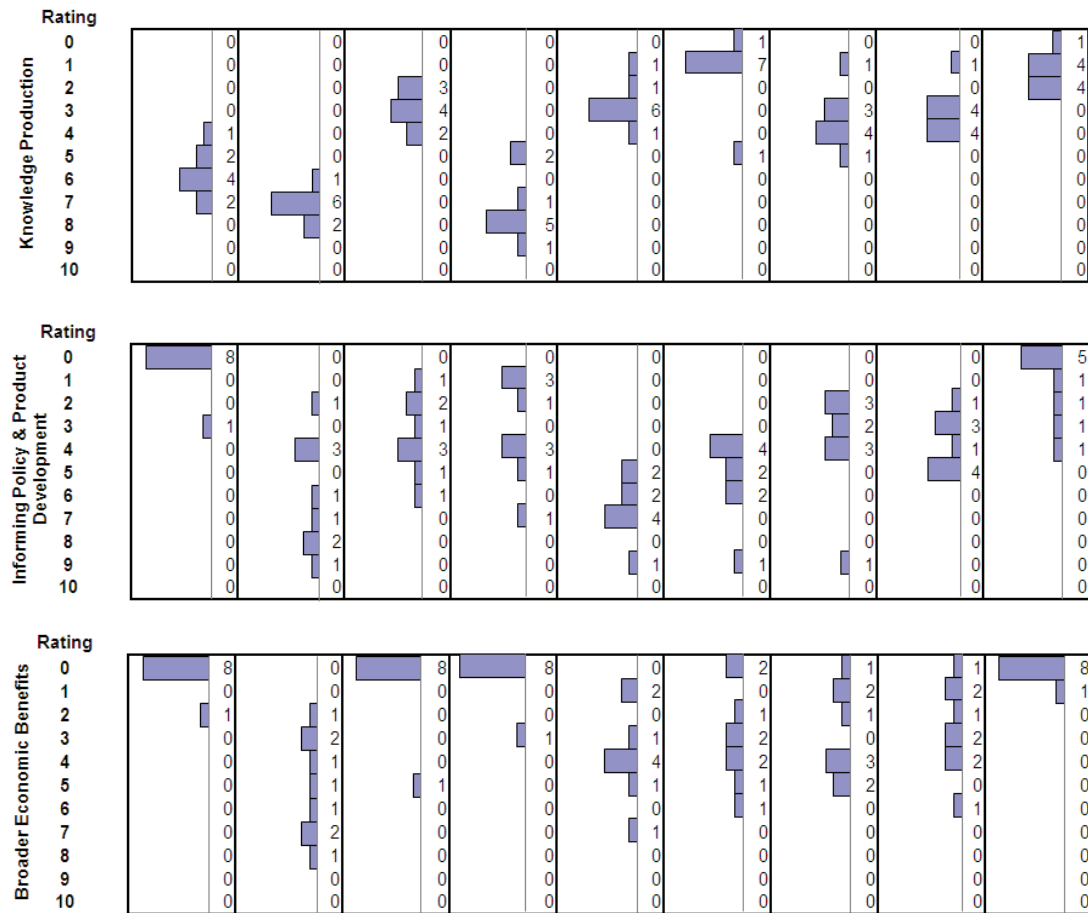
The degree of similarity can also be assessed mathematically by comparing the mean and median ratings that each panel member gave. Only one panel member did not produce a median score of 4, while mean ratings were distributed between 3.3 and 4.3 (see Table 7.1). We repeated the same analysis again after the ratings workshop and found that the behaviour of panel members had become even more similar.

Table 7-1: Summary of the mean and median ratings pre- and post- workshop

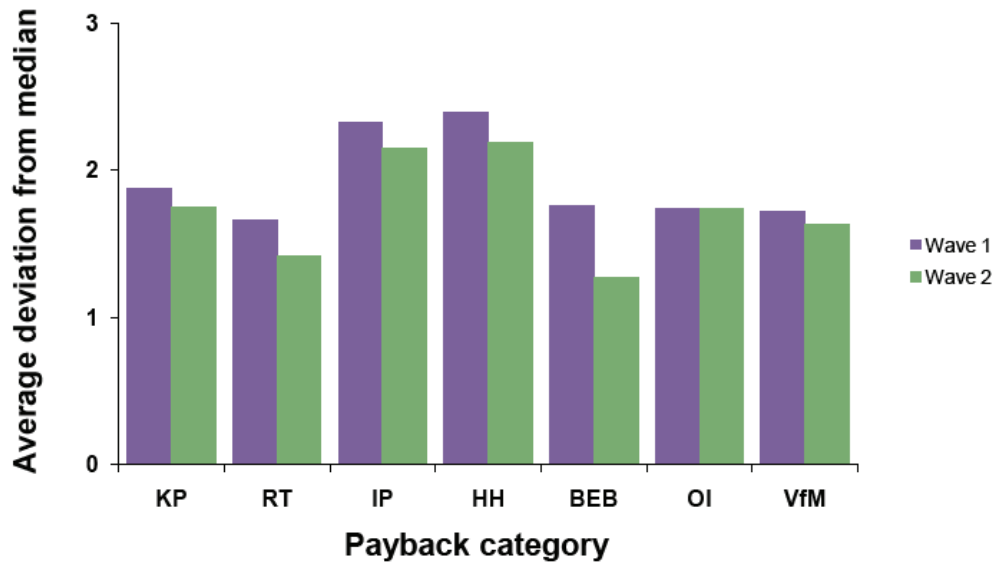
	Pre-workshop		Post-workshop	
	Mean	Median	Mean	Median
Rater 1	3.6	4	3.7	4
Rater 2	3.8	4	3.7	4
Rater 3	3.7	4	3.8	4
Rater 4	4.0	4	4.0	4
Rater 5	3.3	3	3.6	4
Rater 6	4.2	4	4.1	4
Rater 7	4.3	4	4.2	4
Rater 8	4.0	4	3.9	4
Rater 9	4.3	4	4.0	4

After assessing the behaviour of our panel to ensure that members were rating in a similar fashion, we examined the level of agreement for the ratings they produced for each case study in each payback category. To assess this we examined the distributions of ratings and calculated the ADM for each case study in each payback category. Figure 7.4 shows a sample of post-workshop distributions of ratings for the knowledge production, informing policy and product development and broader economic benefit categories. We show these three categories as they demonstrate the range of agreement levels produced.

Figure 7-4: Ratings agreement



ADM represents the average deviation of ratings in a payback category from the median rating in that category: the greater the deviation, the greater the diversity of ratings. Using the ADM measure we could show that agreement in all of the categories improved between the pre-workshop and post-workshop scores, as shown in Figure 7.5, where Wave 1 gives the ratings prior to the workshop and Wave 2 shows the adjusted ratings after the workshop discussions. As discussed previously, we were not attempting to reach consensus among raters, merely to reduce differences in understanding. Although the level of agreement improved, the workshop process changed relatively few of the median ratings for the individual categories within the case studies – 42 of the 203 (21%) changed, with the majority of these changes being a shift of 1 (37 changes, 18% of total ratings). These tests reassured us that our rating process was robust and that there was a similar understanding among raters of level of payback for each case study in each category.

Figure 7-5: Average deviation from the median before and after the workshop

7.3 Confounder analysis

We identified three key variables that might confound our findings: the basic biomedical versus clinical distinction between case studies, the scope of the studies, and studies that produced negative results.

For each of these potential confounders we carried out a sensitivity analysis to test whether they were likely to affect our findings.

7.3.1 Clinical and basic biomedical distinction

Although the concept of distinct categories of basic biomedical and clinical research is almost universally accepted in medical research, it can be very hard to decide exactly how the two different types of research should be defined. There is always likely to be some research that does not sit obviously in either category or which deliberately incorporates elements of both.

As discussed in Chapter 3, during the case study selection process we found that each of our funders had a slightly different idea of the distinction between basic biomedical and clinical research. The definitions we used in the selection of case studies for Project Retrosight were, therefore, a hybrid of our funders' definitions:

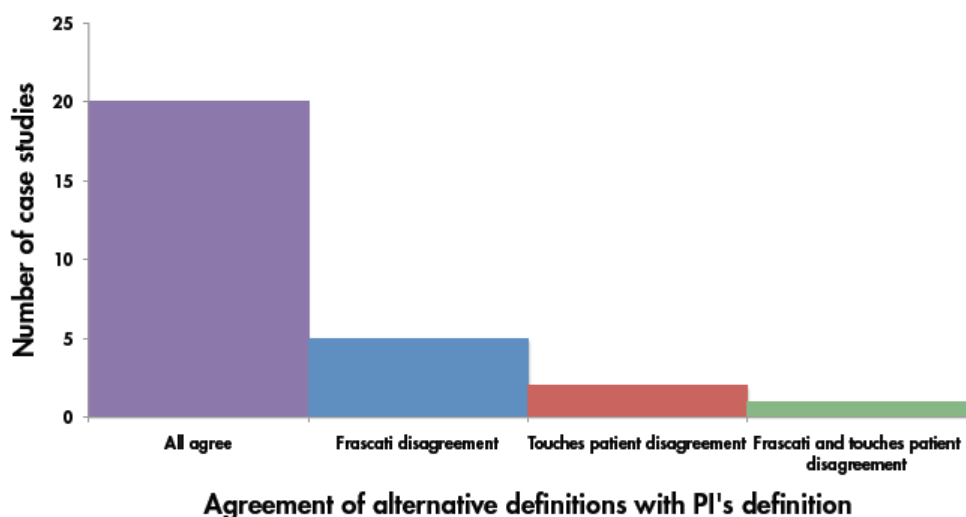
- basic biomedical – focuses on normal or abnormal function at the molecular, cellular, organ or whole body level;
- clinical – focuses on patients, better diagnostics or treatments and increasing quality of life.

These were the definitions we asked PIs to use in classifying their own research for our survey; but because the basic biomedical/clinical distinction is not universally used, we wanted to test whether our observations would be robust against alternative definitions.

We therefore tested the internationally recognised Frascati Manual classification published by the OECD⁵ and a more informal definition of whether or not the research involved direct contact with patients. By applying these two alternative definitions to each case study we were able to gauge the clarity and reliability of the classification system that we initially used in selecting grants. Each definition was applied independently by one member of the research team. The results of the comparison (Figure 7.6) show a generally good agreement between the PIs' definitions and the Frascati Manual, and even closer agreement with the "touches patients" definition.

To examine whether changes to the classification of the case studies affected our observations of factors associated with impact, we tested the effect of "flipping" the case studies where our three sets of definitions had least agreement. The results of this robustness testing are discussed in relation to each factor in the Policy Report in which we discuss our findings (Wooding et al., 2011).

Figure 7-6: A comparison of the alternative research type definitions as applied to the case studies



7.3.2 Case study scope

Nearly all scientific research is undertaken in the context of a body of work being conducted by the researcher over a period of time and in collaboration with a number of colleagues. Such issues of attribution are discussed in the Policy Report (Wooding et al., 2011). One particular aspect of attribution relevant to this project was case study scope; we wanted to test whether the fact that different case studies had different degrees of overlap with other research being carried out by the group affected our assessments of level of impact.

⁵ In order to test agreement with our basic biomedical/clinical definition we treated Frascati basic research as equivalent to basic biomedical and Frascati applied research as equivalent to clinical.

While some studies of the impact of research attempt to examine the whole portfolio of research over a given time period, we wanted to focus on specific research grants in order to measure the payback from research funding. Hence one potential confounding factor for Project Retrosight was the actual scope of the research in our sample case studies.

Our main concern was that in at least some cases the impacts we identified would, in reality, be the result of a longer or broader body of work and of multiple sources of funding. We therefore had to find a way of isolating the impacts generated by the specific grant we were studying from the impact generated by the PIs' previous work, their later research and any body of work or research funding that happened concurrently.

This was done by analysing the case study summaries, reviewing the appropriate passages of case study text and carrying out two brief analyses of scope based on the focus of the research. We then combined our observations and used them to grade the scope of each case study.

We assessed the level to which each case study included impacts derived from the PIs' previous work, wider concurrent research/funding and subsequent work, and graded each on a scale of three: low (assigned a value of 0); moderate (assigned a value of 1); and high (assigned a value of 2).

Generally speaking, neither the existence of previous or contemporaneous research, nor evidence that the PI was receiving concurrent funding, led automatically to a high score for scope. Rather, the scores indicating a high level of scope were given when the impacts described could to some extent be associated with previous or wider research. Particular problems arose in relation to rating the scale for contemporaneous research because it was sometimes very difficult to tell if a wide range of related contemporaneous studies being conducted by the PI's team or lab was contributing to the impacts described in the case study. Sometimes comments in the detailed case studies give clues: for example, one PI stated that the grant from the NHF of Australia covered just a quarter of his project; another stated that there was no overlap in terms of funding between his existing projects and the project that formed the subject of the case study.

It was decided to take the combined score from the two scales for previous research and contemporaneous research/funding and use this to decide whether scope was a confounder. The third scale (impact attributed to follow-on research) was discounted because it seemed logical that the impacts from follow-on research can and should to some extent be credited to the original project. We saw no relationship between the extent of overlap with previous or contemporaneous work and academic or wider impact. This reassured us that scope was not a confounding factor.

7.3.3 **Negative or null findings**

We classified five of our case studies as reporting negative or null findings. The concept of negative or null results has various definitions⁶; we use it to refer to studies that fail to confirm or disprove a hypothesis, or those that disprove a hypothesis and hence preclude further work (for example, by demonstrating the absence of a link between two systems).

⁶ See the editorial guidelines of *Journal of Negative Results in Biomedicine* and *Journal of Negative Results (Ecology and Evolutionary Biology)*.

As we expected negative or null findings to be independent of other factors that might influence impact, we were concerned that a coincidental association between this and a second factor could lead us to make inaccurate observations. We therefore tested the distribution of case studies with each factor associated with impact both before and after excluding case studies with negative or null findings – the results of this analysis are discussed in the Policy Report (Wooding et al., 2011).

We were concerned about negative findings because the available literature suggests that they are less likely to be published than positive findings; this suggests that they will be associated with lower impact (Gould, 1993). We also think it is likely that our methodology may underestimate the impact of studies with negative findings, as it is likely to be harder to evaluate the absence of an impact – an area of research that is not pursued, saving money for more productive research – than an impact that did materialise.

The final stage of the project was to carry out a cross-case analysis of the case studies to learn more about the process of cardiovascular and stroke research and to identify factors that appeared to be associated with the successful translation of research outputs into impacts on policy, improvements in health care and wider economic benefits.

8.1 Deriving impact categories

Once we had quantified the payback from each case study, in each payback category, we examined the median ratings to look for patterns. We were concerned that identifying factors associated with high impact in all five payback categories would be unmanageable and so, if possible, we wanted to reduce the number of payback measures. We found that grouping the first two payback categories and the last three allowed us to produce two mutually exclusive groupings that also grouped the payback categories with the highest correlation (Figure 8.1).

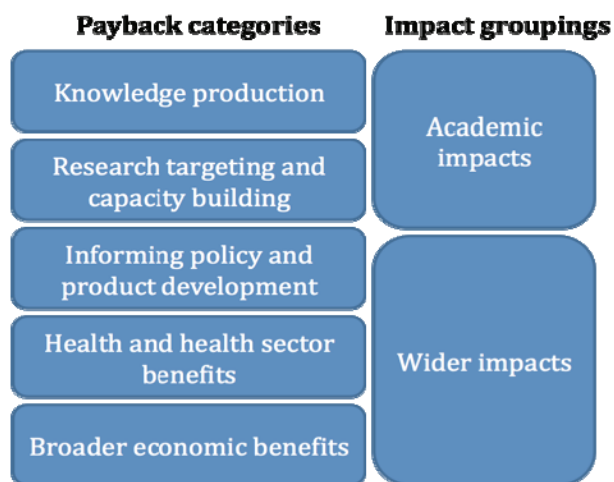
Table 8-1: The correlations between ratings in each payback category

	Knowledge production	Research targeting & capacity building	Informing policy & product development	Health & health sector benefits	Broader economic benefits
KP	1	0.424	0.112	0.106	0.036
RT	0.424	1	0.286	0.334	0.187
IPPD	0.112	0.286	1	0.748	0.441
HHSB	0.106	0.334	0.748	1	0.52
BEB	0.036	0.187	0.441	0.52	1

Coefficients in **bold** are significant at a 0.01 level
Spearman's rho and 2-tailed significance; n = 261 (29 projects x 9 raters)

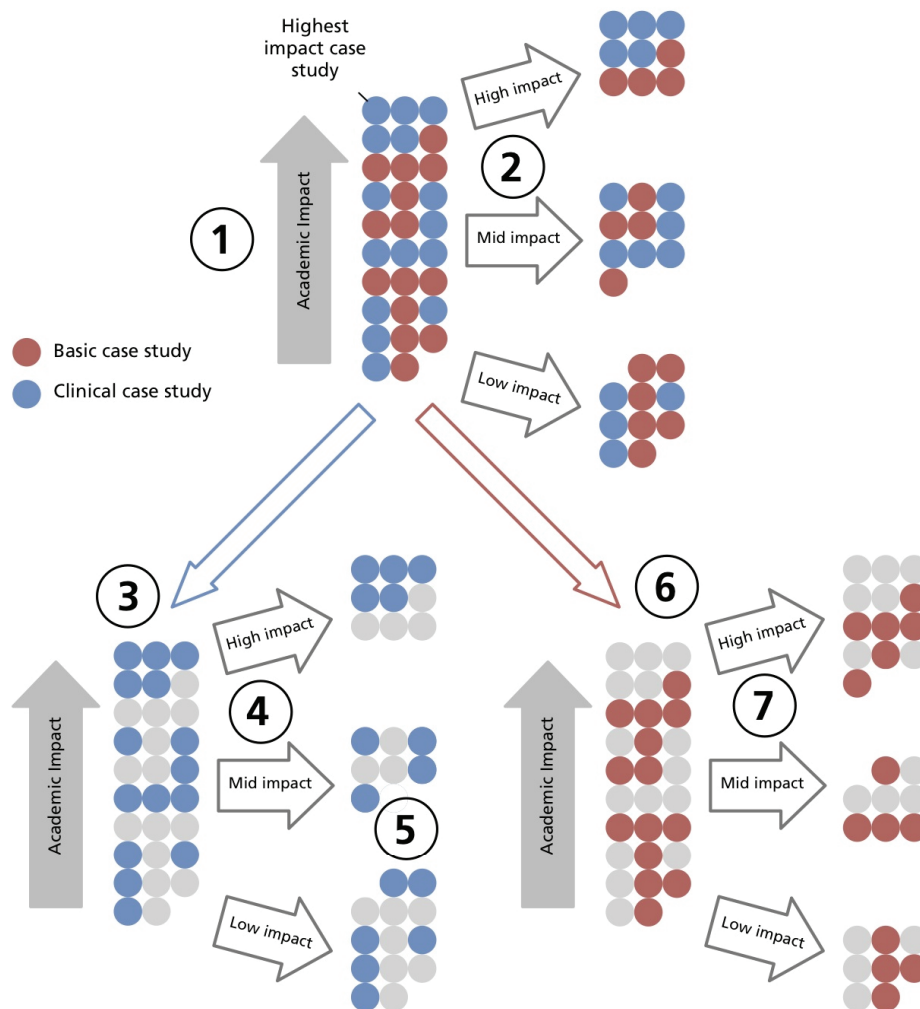
We grouped the payback categories in this fashion to produce two impact categories: academic impact, encompassing knowledge production and research targeting and capacity building; and wider impact, encompassing informing policy and product development, health and health sector benefits, and broader economic benefits (Figure 8.1). These groupings are also intuitively attractive, in that our first category covers impacts affecting science and/or the research system, while our second represents impacts on broader society.

Figure 8-1: Payback categories and impact groupings



In order to examine similarities and differences between high- and low-impact research, we first needed to divide our case studies into these groups. For each of the academic and wider impact categories we sorted the complete list of case studies by impact and alternately took one case study from the top (high-impact group) and one from the bottom (low-impact group) until the middle part of the scale between the two groups contained ten case studies. Case studies with tied scores were included or excluded to ensure that the gap was at least ten. This resulted in the complete list being divided as closely as possible into three equal groups (see Figure 8.2 for diagrammatic representation of the process). As there was some uncertainty in our measures of impact we used a middle impact group to ensure that we had a clear distinction between our high and low impact cases, rather than two overlapping groups.

Figure 8-2: A diagrammatic overview of the analysis process to identify factors associated with impact



To prepare our groups of case studies for analysis to identify factors associated with academic impact: we first sorted all the case studies (basic and clinical together) in ascending order of academic impact (1) then split them into three roughly equal groups: high impact, mid impact and low impact (2). See methodology report for exact details of splitting methodology.

To look in more detail at the characteristics of basic biomedical research we took only the basic case studies (3) (clinical case studies shown faded to grey) and split them into high, mid and low impact groups (4). When we were considering only the basic case studies the low/mid split point fell in a slightly different place. This means two basic case studies are mid impact when considering all the case studies but low impact when considered with just the basic case studies (5). We then carried out a similar process for the clinical case studies by themselves (6&7) (basic case studies shown faded to grey). On this occasion both split points differed.

This sorting process gave us six groups of case studies: high, mid and low academic impact; alongside high, mid and low wider impact. There was overlap between the case studies which appeared in the high academic impact group and those which appeared in the high wider impact group, but there were also differences, as we would expect from the low level of correlation between early and late payback categories. There were similar differences between the mid- and low-impact groups. These overlaps are shown in Table 8.2. We also did a similar grouping of impact after splitting case studies into basic biomedical and clinical categories, as shown in the lower half of Figure 8.2. Again, although the groupings are similar, some case studies end up in different groups; for example, some basic

biomedical case studies classified as mid wider impact move to the high wider impact category when only basic biomedical case studies are considered (see Table 8.3).

Table 8-2: Overlap between high, mid and low groupings for academic and wider impact

	High academic	Mid academic	Low academic
High wider	4	4	1
Mid wider	3	2	5
Low wider	2	4	4

Table 8-3: Overlap between the high, mid and low groups for all cases studies versus the groupings for basic and clinical case studies separately

Academic impact	High basic biomedical	Mid basic biomedical	Low basic biomedical	High clinical	Mid clinical	Low clinical
High all	5			4		
Mid all		4	2	2	2	
Low all			4		2	4

Wider impact	High basic biomedical	Mid basic biomedical	Low basic biomedical	High clinical	Mid clinical	Low clinical
High all	5			5	1	
Mid all		4			4	2
Low all			6			2

8.2 Identifying factors associated with research impact

In this stage of the analysis we aimed to identify factors common in high-impact research but not shared by (or at least uncommon in) low-impact research. As explained above, we considered academic impact and wider impact separately because the two were not well correlated – a case study could have high academic impact but low wider impact, or vice versa.

Four members of the research team (MB, JG, AP and SW) each took one of four impact groups (academic/high impact, academic/low impact, wider/high impact, wider/low impact) and independently examined four of the relevant case studies in their category to look for common factors. The two researchers looking at high-impact case studies then compared their findings and identified emerging factors associated with impact that they had both observed. The two researchers examining low-impact case studies followed the same procedure.

This exercise was followed by a discussion of the factors identified, involving the whole research team; during this discussion the team assembled a core set of factors on which to focus. These were classified as factors of high importance and factors of potential interest that lacked further evidence and were not strong enough to be the main focus. Figure 8.3 shows the complete list of factors compiled at this stage.

Figure 8-3. Factors coded for in the 29 case studies. Bold italics indicate factors analysed and discussed in the Policy Report (Wooding et al., 2011)

- **Clinical motivation**
- **Basic biomedical research in a clinical setting**
- **Strategic thinking**
- **Collaboration with other researchers**
 - **Collaboration on resources**
 - **Collaboration on design**
 - **International collaboration facilitating research**
 - **International collaboration facilitating impact**
 - **Other**
- **Engagement with practitioners and patients**
- **Collaboration with industry**
- **Negative or null findings**
- **Initial rejection of application**
- **Basic biomedical and clinical research perspectives are combined**
 - **In the PI**
 - **In the team**
- **Research only has basic biomedical perspective**
- **Research only has clinical perspective**
- **Career stage of PI**
- **Field transition**
- **New method in the field**
- **Confronting prejudice or entrenched viewpoints**
- **Policy collaboration / engagement with policymakers**
- **Fashionable area of science**
- **Initial cutting of grant resulted in follow-on/extension of funding**
- **Lack of standard commercialisable outcomes**
- **PI has long term record in field**
- **Follow on research produced negative findings**
- **Senior champion for the PI's work**
- **Use of animals in research**

Each person then reread their initial four case studies, coding for the agreed factors, took the opposite end of the other impact category (for example, switched from academic/low impact to wider/high impact) and coded the remaining cases in that category.

The coding at this stage was done in NVivo, a widely used software package for qualitative analysis⁷. It allows text relevant to each factor to be highlighted and coded to the factor, and for the coding of all four researchers to be combined in one file.

Each factor was analysed quantitatively and qualitatively. Quantitative analysis was carried out in NVivo to determine the number of case studies in which each factor occurred. The frequency at which these factors occurred was compared in the high- and low-impact groups defined previously, both for academic and for wider impact categories.

In addition to this overall comparison, each of these ranking lists was split out into basic biomedical and clinical research (so there were four separate ranking lists: basic biomedical/academic, clinical/academic, basic biomedical/wider and clinical/wider). Again, high- and low-impact groups were determined from these scales. The same method of defining high- and low-impact groups was used again, with the gap between high and low groups set at four case studies rather than ten due to the smaller number of cases in each list.

The following list of analyses was conducted in NVivo (in each instance, the frequency of occurrence of each factor was compared).

- High- vs low-impact academic across all case studies
- High- vs low-impact academic basic biomedical case studies
- High- vs low-impact academic clinical case studies
- High- vs low-impact wider across all case studies
- High- vs low-impact wider basic biomedical case studies
- High- vs low-impact wider clinical case studies
- High- vs low-impact in both academic and wider across all case studies
- High- vs low-impact in both academic and wider basic biomedical case studies
- High- vs low-impact in both academic and wider clinical case studies

The last three were not used in subsequent interpretation because of the small numbers of case studies with high/low scores in both impact categories. The project team met to discuss the analysis and identify emerging findings.

Qualitative analysis was done by assigning each factor to a coder who reviewed all the mentions of that theme to ensure consistency of interpretation. Resource limitations meant

⁷ We used NVivo 8, which has now been updated; QSR International. *NVivo software*. http://www.qsrinternational.com/products_nvivo.aspx [Accessed 16 December 2010]

that we could not double-code cases, but by coding first by case study and then by factor we hope to have removed the majority of inconsistencies.

Key messages were drafted and supporting evidence identified for discussion at the Emerging Findings workshop.

8.3 Emerging Findings workshop

To test and review our emerging conclusions and help us develop policy-relevant findings we convened an international workshop in May 2010 with a range of stakeholders from the biomedical and cardiovascular research community (as shown in Table 8.4).

Table 8-4: Emerging Findings workshop participants

Name	Affiliation
Liz Allen	Wellcome Trust
Stefano Bertuzzi	Office of Science Policy, National Institutes of Health, USA
David Cox	Research and Development, Department of Health
Vicki Crossley	Economic and Social Research Council
Lisa Croucher	Arthritis Research UK
Sally Davies	Research and Development, Department of Health
Chris Henshall	University of York
Liam O'Toole	Arthritis Research UK
Jeremy Pearson	British Heart Foundation
Rachel Quinn	The Academy of Medical Sciences

At the workshop we presented the methods, results and emerging findings of the study. The bulk of the workshop was then given over to discussion of the issues raised by each stage of the project. This workshop was extremely valuable in helping refine our thinking and suggesting additional focuses for further iterations of the quantitative and qualitative analysis outlined above. Following the workshop we revisited in detail the evidence behind each of our findings and further investigated the subtleties of the data. We refined our observations based on this analysis and developed the policy implications outlined in our main policy report (Wooding et al., 2011).

REFERENCES

Reference list

- Buxton, M., and S. Hanney. "How can payback from health services research be assessed?" *Journal of Health Services Research and Policy* 1 (1996): 35–43.
- Buxton, M. and S. Hanney. *Assessing payback from Department of Health research and development: second report. Volume 1: the main report. HERG Research Report No. 24*. Uxbridge, UK: HERG, Brunel University, 1997.
- Buxton, M., and S. Hanney. "Evaluating the NHS R&D programme: will the programme give value for money?" *Journal of the Royal Society of Medicine* 91, Suppl 35 (1998): 2–6.
- Buxton, M., and W. Schneider. *Assessing the 'payback' from AHFMR-funded research*. Alberta, Canada: Alberta Heritage Foundation for Medical Research, 1999.
- Crocker, L., and J. Algina. *Introduction to classical and modern test theory*. Orlando, FL: Holt, Rinehart and Winston, 1986.
- Fitch, F., S.J. Bernstein, M. Dolores Aguilar, B. Burnand, J. Ramon LaCalle, P. Lazaro, M. van het Loo, J. McDonnel, J.P. Vader and J.P. Kahan. *The RAND/UCLA appropriateness method user's manual*. Santa Monica, CA: RAND Corporation, MR-1269-DG-XII/RE, 2001.
- Gould, S.J. "Cordelia's dilemma." *Natural History* 102 (1993): 10–16.
- Hanney, S., J. Grant, S. Wooding and M. Buxton. "Proposed methods for reviewing the outcomes of research: the impact of funding by the UK's 'Arthritis Research Campaign'." *Health Research Policy and Systems* 2 (2004): 4.
- Hanney, S., M. Buxton, C. Green, D. Coulson and J. Raftery. "An assessment of the impact of the NHS Health Technology Assessment Programme." *Health Technol Assess* 11 (2007): 53.
- Nason, E., B. Janta, G. Hastings, S. Hanney, M. O'Driscoll and S. Wooding. *Health research – making an impact. The economic and social benefits of HRB-funded research*. Dublin: Health Research Board, 2008.
- Oortwijn, W., S. Hanney, A. Ligtoet, S. Hoorens, S. Wooding, J. Grant et al. "Measuring the impact of health technology assessment in the Netherlands." *International Journal of Technology Assessment in Health Care* 24 (2008): 259–69.
- Pang, T., R. Sadana, S. Hanney, Z.A. Bhutta, A.A. Hyder and J. Simon. "Knowledge for better health – a conceptual framework and foundation for health research systems."

Bulletin of the World Health Organization 81 (2003): 815–20.

Pollitt, A., S. Wooding, S. Hanney, M. Buxton, and J. Grant. *Project Retrosight. Understanding the returns from cardiovascular and stroke research: Case Studies*. Cambridge, UK: RAND Europe, WR-836-RS, 2011.

RAND Europe. <http://www.rand.org/about/standards/> [Date accessed: 16 December 2010]

Wooding, S., S. Hanney, M. Buxton and J. Grant. *The returns from arthritis research. Volume 1: Approach, analysis and recommendations*. Cambridge, UK: RAND Europe, MG-251-ARC, 2004.

Wooding, S., S. Hanney, M. Buxton and J. Grant. “Payback arising from research funding: evaluation of the Arthritis Research Campaign.” *Rheumatology* 44 (2005): 1145–56.

Wooding, S., S. Hanney, A. Pollitt, M. Buxton, and J. Grant. *Project Retrosight. Understanding the returns from cardiovascular and stroke research: Policy Report*. Cambridge, UK: RAND Europe, MG-1079-RS, 2011.

Wooding, S., E. Nason, L. Klautzer, J. Rubin, S.R. Hanney and J. Grant. *Policy and practice impacts of ESRC funded research – a case study of the ESRC Future of Work programme, approach and analysis*. Cambridge, UK: RAND Europe, 2006. http://www.rand.org/pubs/technical_reports/TR435/ [Date accessed: 16 December 2010]

Wooding, S., E. Nason, T. Starkey, S. Hanney and J. Grant. *Mapping the impact: exploring the payback of arthritis research*. Santa Monica, CA: RAND Corporation, 2009.

Yin, R.K. *Case study research: design and methods* (3rd ed.). Thousand Oaks, CA: Sage, 2003.

APPENDICES

Appendix A: PI Survey

Project Retrosight Survey

As a research grant holder in cardiovascular disease in the early 1990s, we would like to ask you to kindly participate in a short survey (around 5 minutes to fill in) for a multi-national research project looking at the impacts of CVD research over long time periods. The research is being supported by a consortium consisting of the following bodies:

English Department of Health

British Heart Foundation

UK Stroke Association

Canadian Institute of Health Research

Heart and Stroke Foundation of Canada

New Zealand National Heart Foundation

Health Research Council of New Zealand

National Heart Foundation Australia.

The survey is a preliminary phase of the research, prior to in-depth case study research of specific grants funded by the supporters shown above.

In order to proceed to the survey, please enter the password you were provided in the recruitment email below:

(the password should be of the form of three characters followed by three numbers)

[*second page*]

Project Retrosight Survey

Our database indicates that you were involved with the following study:

No.	Grantholder	Year	Title	Funder
-----	-------------	------	-------	--------

[grant details]

Can you confirm that this is correct?

No. this is not correct

[if 'No' to correct grants question, then taken to page]

Our database indicates that you were involved with the following study:

Delete? No. Grantholder Year Title Funder

{details editable here}

From here, are taken to section one of the survey.

[If yes – new page]

We would like first to ask about you as a researcher, and in section two we would like to ask you specifically about your research project(s).

Section one:

1. As we are trying to provide a representative sample of researchers, based on the population during the early 1990s, we would like to be able to reflect the gender proportions seen at that time. Since gender is not routinely recorded by research funders, please could you indicate your gender below?

Male [] Female []

2. As of 2007, how many years have you been actively involved in research? (i.e. subtract career breaks, maternity leave, etc).
 - a. Post PhD (if applicable) [*drop down menu of 1-50+*]
 - b. In total [*drop down menu of 1-50+*]

Next

Section two:

We would now like to ask you specifically about the following project:

Grant name. (year)

Please answer the following questions with reference to this project.

3. Was the research funded clinical or basic biomedical according to the following definitions?

Clinical – focuses on patients, on better diagnostics or treatments, and on increasing quality of life.

Basic biomedical – focuses on normal or abnormal function at the molecular, cellular, organ or whole body level.

4. Did the research team involve scientists or approaches drawn from more than one scientific discipline?

It only involved one discipline within health

It encompassed more than one discipline within health

It encompassed more than one discipline from outside health research

5. Was one of the research team a practicing clinician at the time of the research?

Yes

No

6. How many publications (i.e. peer-reviewed articles, editorials, books, chapters and publicly available reports) arose primarily from your project? (please do not include abstracts)

--- Please select ---[*drop down menu* 0, 1–4, 5–9, 10–19, 20+]

7. Have the project findings, methodology or theoretical developments contributed to further research by researchers outside the immediate research team (other than systematic reviews or meta-analyses, conducted by others)?

Yes

No

8. Has participation in this research led to higher degrees (e.g. PhD, MD) for any members of the project team and if so, how many?

None

One

More than one

9. Research findings can be used in policy making (either to change policy or continue with the existing policy) in various ways, across different levels of the system, and with varying degrees of impact. These levels of the health system range from decisions taken at a major level, e.g. national public policy, reimbursement decisions by the Ministry of Health, clinical guidelines from nationally respected bodies; to more minor level, e.g. the policies of a local health care unit or a local group of professionals. What level of decision making did your project impact?

Major

Minor

None

10. Research findings might have made a major or minor degree of impact on particular decision(s). A major impact would include being cited as one of the main pieces of evidence behind the decision. A minor impact would include: evidence of the finding being considered as the policy was developed, some reference to it in a guideline. What was the degree of importance of the project in the decision making?

Major

Minor

None

11. Did the research project lead to any involvement with or a product development by industry?

No involvement

Research collaboration

Licensed patent

Spin out company

Product developed by industry

12. Research findings can influence behaviour of health practitioners (or the public) in various ways (both at different levels of the system and with different degrees of impacts). Levels can range from major (i.e. at a wide national or even international level) to minor (i.e. at a local unit level). At what level of the health system were your research findings applied?

Major

Minor

None

13. Research findings might have made a major or minor degree of impact on the particular practice(s) used in healthcare. A major impact might be identifying a new surgical procedure that is taken on; a minor impact might be modifying a current well-established practice. What was the degree of importance of the project on healthcare practice?

Major

Minor

None

Next

[*endpage*]

Your response has been saved.

Thank you for helping with this research.

If you have any questions on this survey please contact Eddy Nason at RAND Europe by email: nason@rand.org

Appendix B: PI interview protocol

Introduction:

RAND Europe is investigating the pathways upon which CVD research is built. RAND Europe is a non-profit policy research organization, with an interest in research policy. In order to carry out this project we are building a series of case studies around research grants that were awarded in the early 1990s.

RAND Europe works in a number of policy areas. This work is being carried out within our Science and Technology program which has an emphasis on research policy. Our previous work in this area includes: Payback studies of research impact on Arthritis Research Council grants and ESRC grants; a study for the National Audit Office on how Government Departments commission research published as 'Getting the Evidence'; and a study to investigate the views of academics with respect to research assessment for the Robert's Review of Research Assessment.

We are looking at both how the findings produced by the grant were developed and translated; and also, how the support developed the careers of researchers.

You will probably want to quickly talk them through the Payback model at this stage.

You should also emphasize that not all the questions will be relevant to their research project, and indeed we wouldn't expect them all to be.

You shouldn't stick to the protocol as written, it just provides guidance of the areas you should aim to cover. During your desk research you will have identified additional questions that you will want to ask and it's probably best to add these to the protocol.

STAGE 0: Opportunity Identification/Research needs Assessment

1. What was the original impetus for your project? Solely scientific curiosity? The need to fill certain gaps in knowledge? Targeting of a particular disease state? Your own clinical experience?
2. How far was your identification of the research topic influenced by:
 - a. Research you had done before? Funded by whom?
 - b. The research of others? If so how did you hear about this research?
3. How much interaction was involved in determining your choice of research topic? With funders? With peers internationally in a specific research community? With representatives of patient or practitioner groups? Did institutional conditions such as lab space, equipment, availability of researchers affect the research proposal?

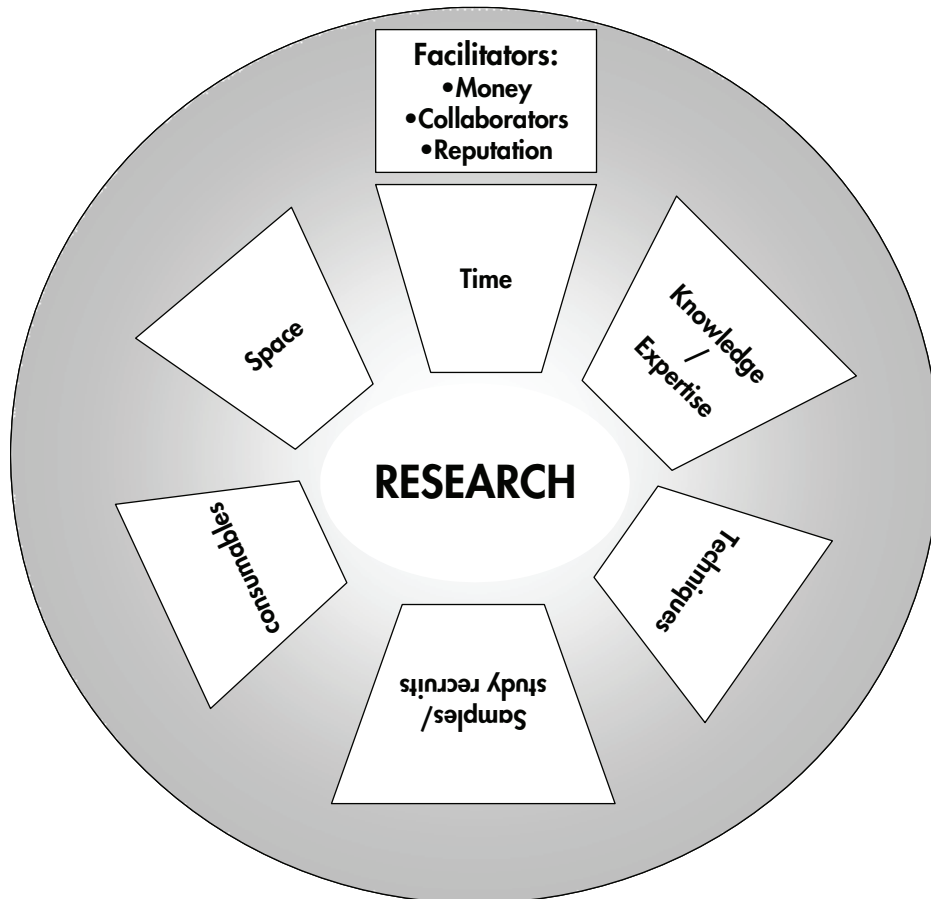
INTERFACE A: Peer review/ Project Specification, Selection and Commissioning

1. Were any changes requested during the peer review/ project selection process?
2. Was there any negotiation involving potential users (users in any sense – maybe clinicians, maybe patients, maybe other scientists) during the project specification or commissioning processes?
3. Was there any involvement of practitioners, or even policy-makers, in the process of project selection?
4. Had your research not been funded by the organisation that did fund it, do you think any other body would have funded the research?

STAGE 1: Inputs to Research

1. Check on cost of project. Were any other resources used on the project, for example where did the overhead costs come from? Was there soft or core funding?
2. What was the institutional setting (hospital, university, research institute) for the research?

3. Who were the main researchers? What was their level of research experience and seniority at that time and in particular had they previously worked in the research area?
4. What inputs of the following were important (provide a copy of the diagram and discuss around the topics):



STAGE 2: Processes

5. Did the methods proposed prove to be appropriate? Were all of the avenues of research suggested in the proposal successful?
6. Was there any interaction with potential users of the research during the research processes?
7. What was your role as PI in the team and the research process? Facilitator? Research director?
8. What was the role of collaborators in the research process (both academic and industrial)?

STAGE 3: Primary Outputs

This section is reliant on the identification of a grant or piece of research from the 1989-1993 period.

1. Identify any publications based on the grant that has been selected and from the subsequent stream of work following the grant.
2. Which publications do you think were most important from this project and why? *(Then reveal citation scores and compare results – based on preliminary Web of Science citation records or bibliometrics if available)*
3. Did this work have any impact on the agenda for your subsequent research?
4. Did it make any impact on the career of any of the research team eg contribute to research training in terms of research degrees or the gaining of additional skills? Enable them to establish themselves in the field? Assist in gaining further? Helping the lead researcher to build a team of researchers?
5. Are you aware of any other researchers who have built on this work or used the methods you developed? What is the role of collaborators in this?
6. Over what type of time-scale do you think your research influenced subsequent research.
7. Did the research spawn a new area of investigation or make a major impact on the approach used in subsequent research? If so would you go so far as to say it led to a paradigm shift in the understanding of a sub-field of arthritis or rheumatology?
8. If the research was clinical or related to AHP were any basic researchers also involved? If so did this influence their attitude to clinical or AHP research?
9. Were any health practitioners involved in assisting with the research, and if so did it have any impact on their attitude towards implementing research findings in general?
10. Did the project play any part in making the existing stock of international knowledge more applicable/acceptable in the UK? Did the project allow work from another field to be applied to arthritis and rheumatology or vice versa?
11. Has the research been included in any formal reviews? *In clinical science this would be a question about systematic reviews, in basic science it is a more general question.*

12. Have you had any impact outside the field of research you are working in?

INTERFACE B: Dissemination

1. Apart from publications, what attempt did you make to disseminate the findings to academic audiences? More widely? Did you work with funders or stakeholders to do this?
2. Did you use specially designed dissemination approaches to particular audiences, for example policy briefs for policymakers? What were the most effective mechanisms for this?
3. What was the role of your networks in dissemination?
4. Did you receive support from funders/employers for dissemination? What form did this take?

STAGE 4: Secondary Outputs

1. Has the research been cited directly in any clinical guideline, audit criteria or similar document from a professional body or public policymaking body at national or local level?
2. Do you know how far the research directly influenced the formulation of any policy, or the realisation that a policy was needed?
3. Has any of the subsequent research by yourself or others that built on the project been cited in any clinical guideline, audit criteria or similar document from a professional body or public policymaking body at national or local level? Do you think this might happen in future?
4. Did the research from your project lead to any patent/licences, was it taken up by industry/has it contributed to any commercial products?
5. If the research has made some impact, what are the key reasons for this? If it has failed to have an impact what are the reasons for this? What **barriers** were there to the research having an impact/being translated? What factors **facilitated** the research having an impact/being translated?

6. Has your research had an impact on teaching for clinicians?
7. Has any advisory role to government, hospitals, industry led to an impact from your research? How did this come about?

Mapping exercise

8. Use a large sheet of paper (flip chart sheet) to map with the PI, the research. Both forwards from the grant under discussion and backwards to understand key inputs to the grant (people, money, expertise – see inputs section). **Error! Reference source not found.** shows an example of what this might look like.

STAGE 5: Applications

1. Have the findings from the research influenced practitioners directly through them reading the articles or hearing a presentation about the research? Has it made any impact on practice through clinical guidelines or policies based either specifically on the research or on other research that built on your research? Has any impact been local, regional, national or international?
2. If the research has been taken up by industry, do you know what level of sales has been achieved by any product to which it contributed?
3. Do you expect any greater take-up of the findings in the future? Where?
4. Has there been an impact on practice through your own clinical work (if you have any)? What has been the knock on effect of that on other clinicians?

STAGE 6: Public Engagement

1. Depending on answers to previous questions about involvement of the public in shaping the research agenda, ask how far there has been any interaction with patients, patient groups or the wider public about the findings and their implication. Has this led to any improvement in the way patients manage their own care or interact with therapy? Or had any impact on public attitudes to medical research?
2. Did engagement with the public/patient groups lead to changes in the researchers' perceptions of these groups?

3. Has there been a change in attitudes in the research community to involvement of the public since the time when this research was conducted?

STAGE 7: Final Outcomes

1. If the research has made impact on policy or practice, or on the behaviour of the public, is there any way of assessing the benefits in terms of: patient health gain? Qualitative improvements in the way the service is delivered that increase patient and/or practitioner satisfaction? Cost savings?
2. If it is possible to assess the potential benefit for one patient, approximately how many patients might be able to benefit from the improved therapy or organisation of the service?
3. If the improved therapy based on the research has resulted in a health gain, will this also result in fewer days lost from work/ decreased benefits payments/ decreased visits to secondary healthcare?
4. If the research has resulted in commercial development is anything known about the amount of employment generated, the level of import substitution, or the revenue generated for the company by the product?

Other general questions

1. Who else should we speak to about your research?
2. Are there other questions we should have asked or things that you want to talk about?
3. Are you happy for us to contact you to follow up on details arising from the case study research?

Appendix C: Non-PI interview protocol

Introduction:

RAND Europe is investigating the pathways upon which CVD research is built. RAND Europe is a non-profit policy research organization, with an interest in research policy. In order to carry out this project for ARC we are building a series of case studies around research grants that were awarded in the early 1990s.

RAND Europe works in a number of policy areas. This work is being carried out within our Science and Technology program which has an emphasis on research policy. Our previous work in this area includes: Payback studies of research impact on Arthritis Research Council grants and ESRC grants; a study for the National Audit Office on how Government Departments commission research published as 'Getting the Evidence'; and a study to investigate the views of academics with respect to research assessment for the Robert's Review of Research Assessment.

We are looking at both how the findings produced by the grant were developed and translated; and also, how the support developed the careers of researchers.

You will probably want to quickly talk them through the Payback model at this stage.

You should also emphasize that not all the questions will be relevant to their research project, and indeed we wouldn't expect them all to be.

You shouldn't stick to the protocol as written, it just provides guidance of the areas you should aim to cover. During your interview with the PI/Lead researcher you will have identified additional questions that you will want to ask and it's probably best to add these to the protocol prior to the interview.

General questions

1. What is your current position? What was your position during the lifetime of the grant?
2. How did you find out about the researcher? The research project in particular?

STAGE 0: Opportunity Identification/Research needs Assessment

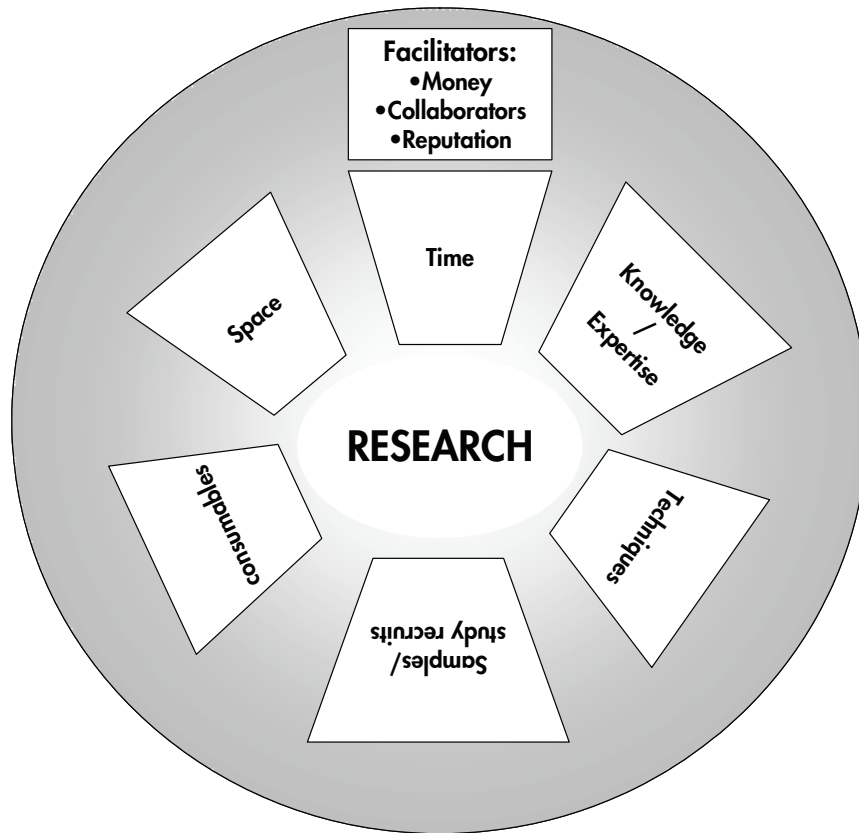
1. Did the project fill certain gaps in knowledge? Was there an urgent need for it (either scientifically or clinically)?
2. Did you have any input into the identification of the research topic/ funding body/ research team identification?
3. Do you know of other research at the time that suggested this project should be performed (either your own research or others)? What was going on in research in this area at the time? Was this a novel project?
4. [For co-researchers] How much interaction was involved in determining your choice of research topic? With funders? With peers internationally in a specific research community? With representatives of patient or practitioner groups? Did institutional conditions such as lab space, equipment, availability of researchers affect the research proposal?

INTERFACE A: Peer review/ Project Specification, Selection and Commissioning

1. Were you involved in any negotiation during the project specification or commissioning processes?
2. Had the research not been funded by the organisation that did fund it, do you think any other body would have funded the research?

STAGE 1: Inputs to Research

1. What inputs of the following were important in helping the project be a success?



(See diagram)

STAGE 2: Processes

1. From what you know, from what you know, were the methods used for the project suitable for the study? Were they groundbreaking/ building on previous research?
2. Was there any interaction with potential users of the research during the research processes?
3. If you were a collaborator in the research process, what was your role? (both academic and industrial)?

STAGE 3: Primary Outputs

1. What specific outputs of the research/researcher do you know? What is your view on their work in general? On this project/subject area?

2. Which publications do you think were most important from this project and why?
3. Did this work have any impact on your own research/ the research field as you know it?
4. [For co-researchers] Did it make any impact on the career of any of the research team eg contribute to research training in terms of research degrees or the gaining of additional skills? Enable them to establish themselves in the field? Assist in gaining further? Helping the lead researcher to build a team of researchers?
5. Are you aware of any other researchers who have built on this work or used the methods?
6. Over what type of time-scale do you think this research influenced subsequent research?
7. Did the research spawn a new area of investigation or make a major impact on the approach used in subsequent research? If so would you go so far as to say it led to a paradigm shift in the understanding of a sub-field of CVD research?
8. What was the relationship between the clinical and basic aspects of the research? If the research was clinical or related to AHP were any basic researchers also involved? If so did this influence their attitude to clinical or AHP research?
9. Do you know of the research being taken on board and used in practice? What facilitated this?
10. Has the research been included in any formal reviews that you know of? *In clinical science this would be a question about systematic reviews, in basic science it is a more general question.*
11. Has the research had any impact outside the immediate field of research? Did the project play any part in making the existing stock of international knowledge more applicable/acceptable in the UK?

INTERFACE B: Dissemination

1. Apart from publications, what other dissemination methods do you know of from this project? (Could be seminars, books, teaching, practice clinics, talks,

presentations, etc...)? Where was this other dissemination (funders/ stakeholders conferences etc.)?

2. Was any particular method of dissemination useful for you? Useful for specific types of information?
3. What was the role of personal networks in this dissemination in your opinion?

STAGE 4: Secondary Outputs

1. Has the research been cited directly in any clinical guideline, audit criteria or similar document from a professional body or public policymaking body at national or local level that you know of?
2. Do you know how far the research directly influenced the formulation of any policy, or the realisation that a policy was needed?
3. Has any of the subsequent research by the researcher, or others that built on the project, been cited in any clinical guideline, audit criteria or similar document from a professional body or public policymaking body at national or local level? Do you think this might happen in future?
4. Do you know of any patent/licences or industry take up/contribution to any commercial products that arose from the research?
5. If the research has made some impact, what are the key reasons for this? If it has failed to have an impact what are the reasons for this? What **barriers** were there to the research having an impact/being translated? What factors **facilitated** the research having an impact/being translated?
6. Has the research project had an impact on teaching for clinicians?
7. What other routes do you know of that could lead to an impact from the research (e.g. any advisory role to government, hospitals, industry)? How would/did this come about?

STAGE 5: Applications

1. Have the findings from the research influenced practitioners directly through them reading the articles or hearing a presentation about the research? Has it made any impact on practice through clinical guidelines or policies based either specifically on the research or on other research that built on your research? Has any impact been local, regional, national or international?
2. If the research has been taken up by industry, do you know what level of sales has been achieved by any product to which it contributed?
3. Do you expect any greater take-up of the findings in the future? Where?
4. Has there been an impact on practice through the clinical work of the researcher involved that you know of? What has been the knock on effect of the research on other clinicians?

STAGE 6: Public Engagement

1. Do you know how far there has been any interaction with patients, patient groups or the wider public about the findings and their implications. Has this led to any improvement in the way patients manage their own care or interact with therapy? Or had any impact on public attitudes to medical research?
2. Has there been a change in attitudes in the research community to involvement of the public since the time when this research was conducted?

STAGE 7: Final Outcomes

1. If the research has made impact on policy or practice, or on the behaviour of the public, is there any way of assessing the benefits in terms of: patient health gain? Qualitative improvements in the way the service is delivered that increase patient and/or practitioner satisfaction? Cost savings?
2. If it is possible to assess the potential benefit for one patient, approximately how many patients might be able to benefit from the improved therapy or organisation of the service?
3. If the improved therapy based on the research has resulted in a health gain, will this also result in fewer days lost from work/ decreased benefits payments/ decreased visits to secondary healthcare?

4. If the research has resulted in commercial development is anything known about the amount of employment generated, the level of import substitution, or the revenue generated for the company by the product?

Other general questions

1. Do you think that there is anyone else we should speak to about the research?
2. Are there other questions we should have asked or things that you want to talk about?
3. Are you happy for us to contact you to follow up on details arising from the case study research?

Appendix D: External peer review materials

Project Retrosight: Guidelines for expert peer reviewers

Thank you for agreeing to assist us in this project. Project Retrosight is a collaborative project that aims to trace the impact of cardiovascular research funded between 1989 and 1994. The project aims to examine all forms of research impact, from building research knowledge and capacity; to broader health and socio-economic impacts. To do this we are using a series of case studies from Australia, Canada and UK. Each case study examines the impact of a particular cardiovascular research project funded by the MRC (UK), CIHR or HSFC (Canada) or Heart Foundation (Australia). Through this analysis we hope to provide a better understanding of the processes involved in achieving impacts and to try to identify factors that might facilitate higher levels of impact.

The project methodology has been used previously to assess the impact of health and medical research in a number of areas, including diabetes⁸ and arthritis⁹. We hope this work will help funding bodies involved in the project, and other policy makers, consider how they might more effectively support health and medical research.

To help ensure that the material we are using in this study is as accurate as possible, we would like you to review one case-study. This review will not require an assessment or critique of the scientific approach or methodology undertaken. Rather we would like you to assess whether you think the case-study provides an accurate account of the research undertaken and its impact in the scientific and broader community.

The focus of the case studies is on the impacts: how the research has influenced further research and healthcare. The intention is to make these case studies available to an

⁸ Hanney, Steve, Iain Frame, Jonathan Grant, Philip Green, and Martin Buxton, From bench to bedside: tracing the payback forwards from basic or early clinical research - a preliminary exercise and proposals for a future study, Uxbridge, Middlesex: Health Economics Research Group, Brunel University, 2003.

⁹ Wooding, Steven, S R Hanney, M J Buxton, and Jonathan Grant, The returns from arthritis research: approach, analysis and recommendations. Volume 1, Cambridge, UK: RAND Europe, 2004.

audience who is unlikely to understand the technical detail of the research, so they have been written with only an overview of the work undertaken.

The case studies are based on relevant documentary evidence, interviews with the lead researchers and interviews with other key informants involved in the research undertaken. The lead researcher involved in the project has also reviewed the case study and they have authorised further use in this study. For each project we are seeking the opinion of two experts: one from the same country as the PI and another from different country to add an international perspective.

In your review of this case study, the key questions we would like you to address are:

- Does the account appear to correctly characterise the findings, conclusions and scientific implications of the research project in question?
- Does the account appear to correctly summarise the impacts of the study?
 - Are impacts claimed that you do not think can reasonably be linked to this research?
 - Are you aware of impacts that might have been missed?
- Does the account seem to you to acknowledge research that was being undertaken elsewhere around the same time and which might justifiably also claim to underpin some of the impacts referred to?
- To your knowledge, were there other research teams addressing very similar questions in parallel and who effectively published the same conclusions in advance of the team in question?

Additionally we would like you to advise us:

- Whether the science involved is adequately described, remembering that the accounts are aimed at lay audiences rather than expert scientists?

We have provided a template for you to outline your responses to these questions as you examine the case study. We would appreciate your honest response and any detail you could provide about any concerns you might have with the content of the case study. If you have reservations about the claims made in the account, please indicate concisely, but as precisely as you can, the basis for your concerns. If there is specific literature (papers, guidelines, commentaries) that would support your argument or assist us in coming to a fair conclusion please cite these if possible. If you have sense that there are important inaccuracies in the account but you feel that the particular issue is outside your expertise, please suggest the name of someone that might better be able to comment.

The comments you provide us will be shared anonymously with the lead researcher but will otherwise be kept in confidence.

Table 1: The standard classification of impacts used in the case studies (with indicative examples)

Payback Category	Example
Knowledge Production	Number of peer-reviewed journal articles High Impact publications
Research targeting and capacity building	PhDs and promotions Additional funding
Informing policy and product development	Influence on clinical guidelines New device developed
Health and health sector benefits	Improved health attributable to research Cost savings to the health sector
Broader economic benefits	Revenue generated by start up companies Employment due to research products (e.g. in start-ups)

PROJECT RETROSIGHT: EXPERT PEER REVIEWER FORM

Reviewer:

Date:

Case Study:

The boxes should expand as you type, do not worry if your comments break across pages

Does the account in this case study appear to characterise accurately the findings, conclusions and scientific implications of the research project in question? If you feel there are any inaccuracies in the account, please provide further detail and relevant references.

Does the account appear to correctly summarise the impacts of the research project?

- *Are there impacts claimed that you do not think can reasonably be linked back to this project?*
- *Are you aware of likely impacts that may have been omitted?*

Does the account included in the case study seem to appropriately acknowledge research that was being undertaken elsewhere at around the

same time and which might justifiably also claim to underpin some of the impacts discussed in the case study?

To your knowledge, were there other research teams addressing very similar questions in parallel with this research project? Is there potential that they may have published the same conclusions in advance of the team in question?

Is the science underpinning the project adequately described? In answering this question, please note that the accounts are aimed at lay audiences rather than researchers with expertise in the area.

Do you have any other comments or suggestions you would like to make about the case study you have reviewed?

Thank you for your time. The comments you provide us will be shared anonymously with the lead researcher but will otherwise be kept in confidence.

Appendix E: Rating instructions

Project Retrosight – Guidance for Scorers

Thank you for agreeing to assist us with the scoring stage of this project. Project Retrosight is a collaborative project that aims to trace the impact of cardiovascular research. We examined research funded between 1989 and 1994. The project aims to examine all forms of research impact, from building research knowledge and capacity; to broader health and socio-economic impacts. To do this we use a series of case studies from Australia, Canada and the UK. Each case study examines the impact of a particular cardiovascular research project funded by the MRC (UK), CIHR or HSFC (Canada) or Heart Foundation (Australia). Through this analysis we hope to provide a better understanding of the processes involved in achieving impacts and to try to identify factors that might facilitate higher levels of impact.

The project methodology has been used previously to assess the impact of health and medical research in a number of areas, including diabetes¹⁰ and arthritis¹¹. We hope this work will help funding bodies involved in the project, and other policy makers, to consider how they might more effectively support health and medical research.

This scoring exercise is a means for us to identify research grants which have had a high impact and those which have had a lower impact across a range of categories. This will help us in trying to identify common factors across case studies with similar levels of impact, and hopefully build a picture of factors which lead to research impact.

To do this, we ask you to score 29 case studies for their level of impact in each of a range of payback categories. Each case study relates to one grant awarded between

¹⁰ Hanney, Steve, Iain Frame, Jonathan Grant, Philip Green, and Martin Buxton, From bench to bedside: tracing the payback forwards from basic or early clinical research - a preliminary exercise and proposals for a future study, Uxbridge, Middlesex: Health Economics Research Group, Brunel University, 2003.

¹¹ Wooding, Steven, S R Hanney, M J Buxton, and Jonathan Grant, The returns from arthritis research: approach, analysis and recommendations. Volume 1, Cambridge, UK: RAND Europe, 2004.

1989 and 1994 in the field of cardiovascular research. We will also ask you to give each case study an overall score and a value for money score. We would like you to score the case studies within the set, and not relative to any wider expectations or experience. By this, we mean that the level of impact achieved should only be scored in comparison to the other case studies presented. Please note that the precise scores are not critical, rather they are a means to establish groups of high and low impact case studies. In order to help do this effectively, we suggest the following approach:

1. Familiarise yourself with the case studies by reading over the summaries and reviewing the other materials (outlined below)
2. When you are ready to start scoring, look through the material relating to a particular category of payback for all case studies
3. Identify any case studies which have no impact in a particular payback category and give them a score of 0
4. Select the case study/ies which has/have had the lowest level of impact in that category (relative to the other case studies) and give it/them a score of 1
5. Select the case study/ies which has/have had the highest level of impact in that category (relative to the other case studies) and give it/them a score of 9
6. Score the other case studies relative to these benchmarks

It is possible to have more than one case study scored at 1 or 9 (or any other score). Again, note that the aim of this exercise is to distinguish high and low impact case studies, so it is not necessary to spend a large amount of time attempting to distinguish between case studies with similar levels of impact.

We would like you to score each case study against the following categories:

- **Knowledge Production:** covers the findings from the project that are contained in peer reviewed articles, books etc.
- **Research Targeting and Capacity Building:** includes the ways in which further research by the team, or others, builds on the project's findings and/or methods; and the research training gained through the project, especially in terms of higher degrees, and contribution to career progression of team members.
- **Informing Policy and Product Development:** includes the ways in which research informs the development of policies, broadly defined to include national legislation, local protocols, clinical guidelines from organisations within the health care system and from professional bodies, public health advice offered by health service or professional bodies etc; and the use of research to inform the development of drugs and devices etc. It can also include decisions to take a drug off the market.

- **Health and Health Sector Benefits:** includes health gains resulting from the adoption of the research findings directly by health care professionals, or by members of the public, or as a result of professionals following clinical guidelines or other policies etc; qualitative improvements in the processes of delivering the services; cost savings; and increases in health equity.
- **Broader Social and Economic Benefits:** includes benefits that can arise from the commercialisation of research in terms of spin-off companies, employment, exports, import substitution etc; and benefits gained from having a healthy workforce
- **Overall score:** your assessment of the overall impact of the work across all payback categories combined according to your opinion of their relative value.
- **Value for Money:** your assessment of the value for money provided to society given the size of grant awarded taking into account your assessment of the overall impact and the resources used. We shall send to you by email a table indicating the size of each grant in a directly comparable currency.

In order to help you score these case studies, we provide a range of materials summarising our findings for each case study as follows:

- **Payback table:** Summarises key outputs of the case study under each of the categories listed above. We anticipate the summary table will be the main input to scoring. This also includes a summary of comments from peer reviewers.
- **Case study summary table:** Summarises the case study narrative, organised by stage of the research process. Each section can be cross referenced back to the full case study text, and gives a fuller explanation of the context than the payback table, with additional background material.
- **Peer review summary table:** Summarises the comments received from peer reviewers on the accuracy and clarity of each case study, with particular focus on comments about the level of impact and degree of leadership in the field.
- **Bibliometrics data:** Key data for each case study presented as a series of charts
- **Coding data:** Summary of the key features of the research and outputs from each case study across 2 of the 5 categories.

In addition, an electronic copy of the full case study text and the full comments from peer review are provided for all case studies. We do not expect you will read all the case studies or peer review comments in full; rather, they can be used for reference and clarification where necessary as a supplement to the summary data provided for scoring.

At the scoring workshop, you will have the opportunity to discuss the scoring process and some of your scoring decisions in order to clarify any

misunderstanding. You will then be able to score the case studies for a second time. However, if you do have any questions or queries, do not hesitate to contact us.

We would be grateful if you could return your completed scoring form (by post or in electronic format) to reach us by 9am GMT on the 1st February. This will allow us to complete our initial analysis and prepare materials for the scoring workshop on the 11th/12th February.

Many thanks for your time and efforts. We hope that you will find the process interesting and enjoyable. If you do have any queries please do not hesitate to contact us at any time.