RAND 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 report is part of the RAND Corporation research report series. RAND reports present research findings and objective analysis that address the challenges facing the public and private sectors. All RAND reports undergo rigorous peer review to ensure high standards for research quality and objectivity.

EUROPE

Mental Health Retrosight

Methods and Methodology Report

Susan Guthrie, Steven Wooding, Alexandra Pollitt,
Harold Alan Pincus, Jonathan Grant

RAND EUROPE

Mental Health Retrosight

Methods and Methodology Report

Susan Guthrie, Steven Wooding, Alexandra Pollitt,
Harold Alan Pincus, Jonathan Grant

Prepared for the Alliance of Mental Health Research Funders

The project described in this report was supported in Canada by the Graham Boeckh Foundation, Alberta Innovates Health Solutions, and the Canadian Institutes of Health Research; in the UK by the National Institute for Health Research; and in the USA by the National Institute of Mental Health.

RAND Europe is an independent, not-for-profit policy research organisation that aims to improve policy and decisionmaking for the public interest though research and analysis. RAND's publications do not necessarily reflect the opinions of its research clients and sponsors.

RAND® is a registered trademark

© Copyright 2013 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 (www.rand.org/publications/permissions.html).

RAND OFFICES

SANTA MONICA, CA • WASHINGTON, DC

PITTSBURGH, PA • NEW ORLEANS, LA • JACKSON, MS • BOSTON, MA

DOHA, QA • CAMBRIDGE, UK • BRUSSELS, BE

www.rand.org • www.rand.org/randeurope

Preface

Mental Health Retrosight was a three-year international project that aimed to investigate the translation and payback from mental health and neuroscience research, with a particular focus on schizophrenia. It looked at the development of research over a 20- to 25-year period in Canada, the USA and the UK.

The project was supported in Canada by the Graham Boeckh Foundation, Alberta Innovates – Health Solutions, and the Canadian Institutes of Health Research; in the UK by the National Institute for Health Research; and in the USA by the National Institute of Mental Health. It was the first project funded through the Science of Science for Mental Health Research Network (SOS for Mental Health), a joint initiative between the Graham Boeckh Foundation and RAND Europe. The network was established as a 'think tank without borders' that would undertake research and analysis into mental health research funding.

This report provides a detailed description of the methodology used in the study. Our intention is that this will allow others to reproduce the methodology in order to validate or challenge our findings, or use the approach in other areas. The report is intended to complement the policy report, which describes the policy-relevant findings of the study. There are also a briefing note on the study and two further volumes containing the case studies of research and perspectives on advances in treatment.

RAND Europe is an independent not-for-profit policy research organisation that aims to improve policy and decision making in the public interest, through research and analysis. RAND Europe's clients include European governments, institutions, NGOs and firms with a need for rigorous, independent and multidisciplinary analysis. This report has been subject to RAND's quality assurance process, which consists of rigorous peer review.

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

Susan Guthrie
RAND Europe
Westbrook Centre
Milton Road
Cambridge CB4 1YG
United Kingdom
Tel. +44 (1223) 353 329
sguthrie@rand.org

Contents

Preface iii
Table of figures vii
Table of tables ix

CHAPTER 1 **Introduction** **1**
1.1 Purpose of this report 2
1.2 Overview of methodology and structure of this report 3

CHAPTER 2 **Context** **7**
2.1 Payback Framework 7
2.2 The Retrosight methodology 9
2.3 Adapting the Retrosight approach for this project 10

CHAPTER 3 **Case study selection** **11**
3.1 Unit of analysis: developing the concept of research clouds 11
3.2 Aims of the case study selection process 12
3.3 Bibliometric methodology 13
3.4 Strengths and weaknesses of the approach 15

CHAPTER 4 **Perspectives selection** **17**
4.1 Overview of selection 17
4.2 Delphi-like survey 19
4.3 Review of national clinical guidelines 21
4.4 Combining the Delphi-like survey and guideline review 22
4.5 Strengths and weaknesses of the approach 23

CHAPTER 5 **Building the case studies** **25**
5.1 Structure of case studies and the Payback Framework 25
5.2 Data collection 29
5.3 Validation 33
5.4 Strengths and weaknesses of the approach 35

CHAPTER 6 **Perspectives** **37**
6.1 Structure of the perspectives and approach 37
6.2 Data collection 39

6.3 Validation 40
6.4 Strengths and weaknesses of the approach 40

CHAPTER 7 **Analysis 41**
7.1 Coding the case studies for factors likely to influence impact 42
7.2 Scoring case studies for impact 42
7.3 Combining the data: a statistically guided narrative analysis approach 46
7.4 Examination of classification and possible confounding characteristics 48
7.5 Analysis of perspectives 53
7.6 Strengths and weaknesses of the approach 54

References 57

Appendix A: Details of bibliometric approach for selecting case studies through highly cited papers 59

Appendix B: Interview protocol for case study interviews 69

Appendix C: Codebook 75

Appendix D: The rating workshop 81

Table of figures

Figure 1-1 Methodological approach 2
Figure 1-2 Mental Health Retrosight project stages 3
Figure 2-1 Logic model used in the Payback Framework 8
Figure 3-1 A diagrammatic representation of research clouds 12
Figure 3-2 Case study selection by country and research type 15
Figure 3-3 Case study refusals, and instances where we were unable to contact the researcher 16
Figure 4-1 Breakdown of respondents by occupation 21
Figure 5-1 The Payback Framework's logic model of the research-funding process. 26
Figure 5-2 Histogram showing the number of papers in the research clouds 31
Figure 5-3 Example of bibliometric analysis for forward-tracing case studies 33
Figure 5-4 Number of reviewers per case study 34
Figure 5-5 Summary of reviewer comments on case studies 35
Figure 6-1 Structure of perspectives 38
Figure 7-1 Outline of analysis phase 41
Figure 7-2 Composition of scoring panel 43
Figure 7-3 Statistically guided narrative analysis approach: overview 46
Figure 7-4 The distribution of impact scores by case study across the payback categories for the three classifications examined 51
Figure 7-5 Overall network for pairwise resource comparisons of the case studies 52
Figure 7-6 The mean impact scores of 'Biological' case studies for different types of research across the five impact categories 53
Figure A-1 Schema of process to identify hot research topics using bibliometrics 59
Figure A-2 Case study selection by country and research type 68
Figure D-1 Case studies listed by Average Deviation from the Mean in each category, with selection criteria overlaid 82

Table of tables

Table 4-1 Rate of response by country 20
Table 4-2 List of interventions/treatments for the second round of the Delphi-like survey 21
Table 7-1 Definitions of different types of research 49
Table 7-2 Classification of case studies using different definitions of basic and applied research 50
Table 7-3 Research domain definitions 53
Table A-1 Citation boundaries for different research levels 62
Table A-2 Distribution of papers across countries 62
Table A-3 Distribution of papers by research level and country of corresponding author 63
Table A-4 Distribution of interventional research by type and country of corresponding author 67
Table D-1 Correlation between scorers after workshop 83
Table D-2 Descriptive statistics for each scorer (after workshop) 83
Table D-3 Changes in median score as a result of workshop for all scorers providing scores both before and after workshop 84
Table D-4 Level of correlation between categories using a two-tailed Spearman's Rho. 85

CHAPTER 1 Introduction

There is a common perception that mental health research in recent decades has not translated well from bench to bedside (see, e.g., Insel, 2009). This project aimed at identifying where translation had occurred, and where it had to understand why and how. With a particular focus on schizophrenia, the study aimed to:

- identify the long-term 'payback'[1] from mental health research;
- identify factors that are associated with the successful translation of research;
- provide insights that will inform future funding policy.

As such this project forms part of the growing field of the science of science and seeks to identify success and to understand how that success occurred and how it might be replicated (Grant and Wooding, 2010). It is no longer enough to campaign for more funding for science. At a conceptual level, we need to understand what factors lead to research success. For example, what kind of science, what kind of scientists, what kind of settings and what kind of funding mechanisms are most effective in promoting discovery and translating the most promising research findings from bench to bedside? This project aimed to explore these questions in the arena of mental health.

Mental Health Retrosight looked at the development of research over a 20- to 25-year period in Canada, the USA and the UK. Building on Retrosight methodology previously applied by RAND Europe in the fields of arthritis research (Wooding et al., 2004) and cardiovascular research (Wooding et al., 2011), case studies of research carried out in the late 1980s explored how the findings, methods and ideas from the research were built on and developed up to the present day. A complementary stream of work within the project looked backwards in time[2] from treatment advances made more recently, exploring the research that contributed to their development, as well as other barriers to and facilitators

[1] We define 'payback' to mean the academic, health, social and economic benefits of research as commonly used in research impact assessment and defined by Buxton and Hanney (1996) in the development of the 'payback model'. We do not imply an estimate of the return on investment, as sometimes used in economic analysis.

[2] We have attempted to trace back as far as the key pieces of research that contributed to each advance as far as this is possible, and as a result how far backwards in time we have explored differs between the perspectives.

for the adoption of the advances into practice in a series of 'perspectives'. Figure 1-1 illustrates these two overlapping approaches.

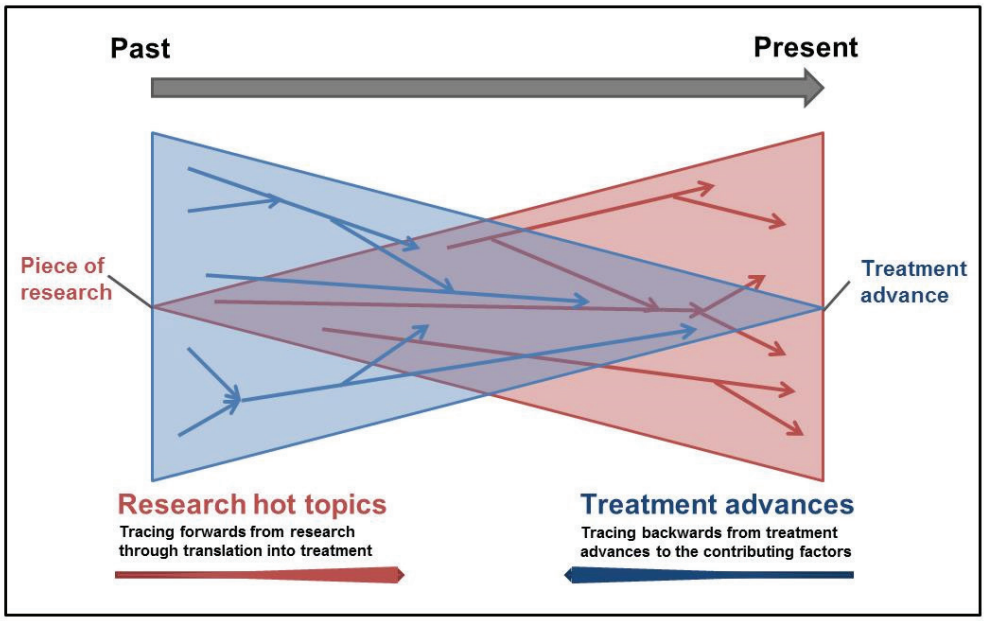


Figure 1-1 Methodological approach

By combining these two methodologies we aimed to analyse as thoroughly as possible the perceived problem that relatively few improvements in mental health care – and schizophrenia care in particular – have emerged; they allow us to investigate both the impacts (or lack thereof) from research in 'hot topic' areas funded in the late 1980s (which may well not have brought improvements in treatment) and the contributory research and contextual factors that *have* led to improved treatments in recent years.

Through these overlapping backward-tracing and forward-tracing streams, the project aimed to identify the payback from research and the factors that are associated with successful translation, and to combine these to provide insights to inform future policy.

1.1 Purpose of this report

This report aims to give a detailed description of the methodology used in this study, to allow a full analysis and critique of our approach, and to enable others to reproduce and build on the methodology used.

1.2 Overview of methodology and structure of this report

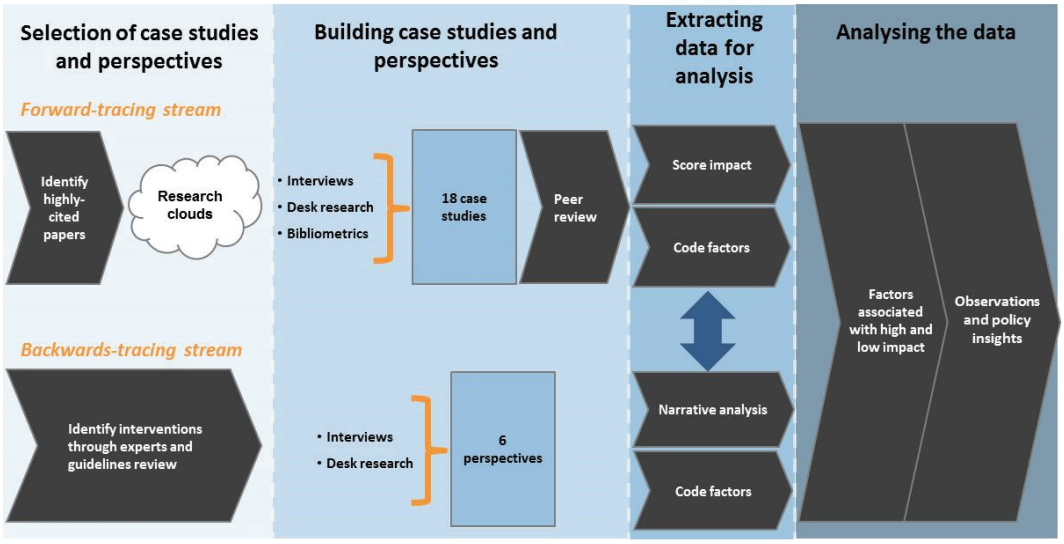


Figure 1-2 Mental Health Retrosight project stages

Figure 1-2 provides an overview of the methodology and the structure of this report. The project broadly consisted of two streams as described above: a forward-tracing stream which started from hot research topics and traced them forward through case studies to look at the impacts that resulted and the processes by which they came about; and a backward-tracing stream which started from a treatment advance and traced them backwards through perspectives to look at the factors, including the research, that contributed to their development.

The first stage of the project was the selection and scoping of the case studies and perspectives. Case studies were selected on the basis of a bibliometrics exercise, with the intention of covering hot topics in research around 20 years ago. The perspectives were identified through a Delphi-like survey of experts in the field and a review of clinical guidelines, and by drawing on the advice of our advisory panel. The aim was to identify some of the key advances in treatment in the field that are affecting current practice.

The next stage of the project was to collect the relevant data. Both case studies and perspectives relied primarily on interviews with the relevant researchers and practitioners and on desk research, covering the academic literature, practice guidelines and wider relevant materials. The case studies were further informed by bibliometrics.

The data-validation step for both the case studies and perspectives consisted of review of the drafts by the researchers contacted for interviews, and for the case studies additional peer review by relevant experts in the field who were not involved directly in the case study. The primary aim of the peer-review step was to validate the extent and nature of the impacts of the work described and to ensure that the appropriate acknowledgement was given to other related work in the field.

The next stage was to convert the case-study narratives in the forward-tracing stream into a manageable dataset for analysis. This consisted of two elements. The first was a ranking exercise in which the case studies were compared in terms of the level of impact in a range of categories corresponding to the categories in the Payback Framework that provides the theoretical framework for the forward-tracing case studies. This ranking was used to identify 'high impact' and 'low impact' case studies so factors associated with success could be identified. The ranking was carried out by a panel of experts including researchers, practitioners, policymakers and service-user representatives. The other element of the process used a coding approach in which the case studies were coded to identify which factors were present in the case studies. An initial list of factors for analysis was identified through the findings of previous studies and suggestions from the team, and this preliminary codebook was developed and built upon iteratively throughout the coding process. The outcomes of these two steps were then synthesised and a statistically guided narrative analysis process used to identify initial findings and generate policy provocations. These findings were presented at an emerging-findings workshop attended by a mix of researchers, research funders and patient representatives from across the three study countries, where the findings were discussed and refined to develop actionable policy insights. These were then tested and built upon through a narrative analysis of the perspectives from the backward-tracing stream.

These stages are described in the subsequent chapters of the report, along with contextual information about previous studies, as follows.

Chapter 2: Context
Sets the scene for the work and explains how it fits into a series of studies that have been carried out using similar methodologies. The chapter also outlines the Payback Framework that we used to structure the case studies in the project.

Chapter 3: Case study selection
Explains in detail how we went about selecting the case studies, first by identifying highly cited papers and then by allocating these into a selection matrix from which we randomly selected case study candidates.

Chapter 4: Perspectives selection
Covers the selection of the subjects for the perspectives using a combination of an on-line survey, a review of guidelines and the input of our expert advisors.

Chapter 5: Building the case studies
Lays out the detailed methodology we used to build the case studies – and includes a discussion of the detailed case study structure and the data collection, which involved interviews, bibliometric analysis, document review and a number of steps designed to ensure the accuracy of the work.

Chapter 6: Building the perspectives
Discusses the methodology used to construct the perspectives – this is somewhat similar to the previous chapter on case studies, although the recruitment of interviewees was somewhat different, as was the review process.

Chapter 7: Analysis

Details how the case studies and perspectives were analysed to derive the findings of the study. For the case studies this included scoring impact with an international panel and coding the case studies for mentions of particular factors that might influence impact. The chapter also presents some supplementary analysis that was carried out on the case studies to characterise better our sample and the detailed additional analyses around particular findings – such as the difference in impact between basic and clinical research.

Each chapter contains a final section discussing the strengths and weaknesses of the approaches taken.

CHAPTER 2 Context

There is an increasing interest in understanding and evaluating the impact of research. This is driven by a range of factors, from the need to demonstrate accountability to taxpayers and other donors for the money spent on research and the desire to demonstrate the benefits research can bring, to the need to improve the allocation of a limited pot of research funding. However, one of the most crucial reasons for understanding the process by which research takes place and leads to wider impacts is to improve on the research process itself and maximise its impacts. This type of analysis requires understanding of detailed contextual factors and the process of the translation of research. Methods such as case studies are one of the best ways in which to explore issues of this kind. Like other research in this area, this study therefore uses case studies to explore the translation of research from bench to bedside. Over the last 20 years the most widely used approach to case studies of this nature in the biomedical sciences has been the Payback Framework, and this was also adopted here.

2.1 Payback Framework

The Payback Framework is a framework for the analysis of research translation initially developed by the Health Economics Research Group at Brunel University (Buxton and Hanney, 1996) to examine health services research, and subsequently developed in collaboration with RAND Europe and other groups for wider applications. The payback model has been applied in several contexts including early clinical research and basic research as well as health services, policy research and, more recently, social science and arts and humanities research (e.g. Hanney et al., 2006; Wooding et al., 2005; Levitt et al., 2010; Panel on Return on Investment in Health Research, 2009). The framework has seen widespread adoption outside the UK, including the CIHR Impact Assessment Framework, the Netherlands, Ireland and Hong Kong (Oortwijn et al., 2008; Nason et al., 2008; Kwan et al., 2007). In Canada the Canadian Academy of Health Sciences panel on return on investment in health research based their recommendations for all users and supporters of health research in Canada on the Payback Framework (Panel on Return on Investment in Health Research, 2009).

The Payback Framework consists of two elements: a classification system to capture and categorise the outputs and outcomes of research, and a logic model which helps to understand and break down the research process. As such it helps not only to evaluate the

range and nature of outputs from research but also to conceptualise the process through which these outputs are generated. The logic model (Figure 2-1) has been used to develop the case study narrative in this and previous studies and is described in more detail in Chapter 1. The classification system has been used to categorise the impacts identified throughout the study and forms the basis of the analysis of impact in this study.

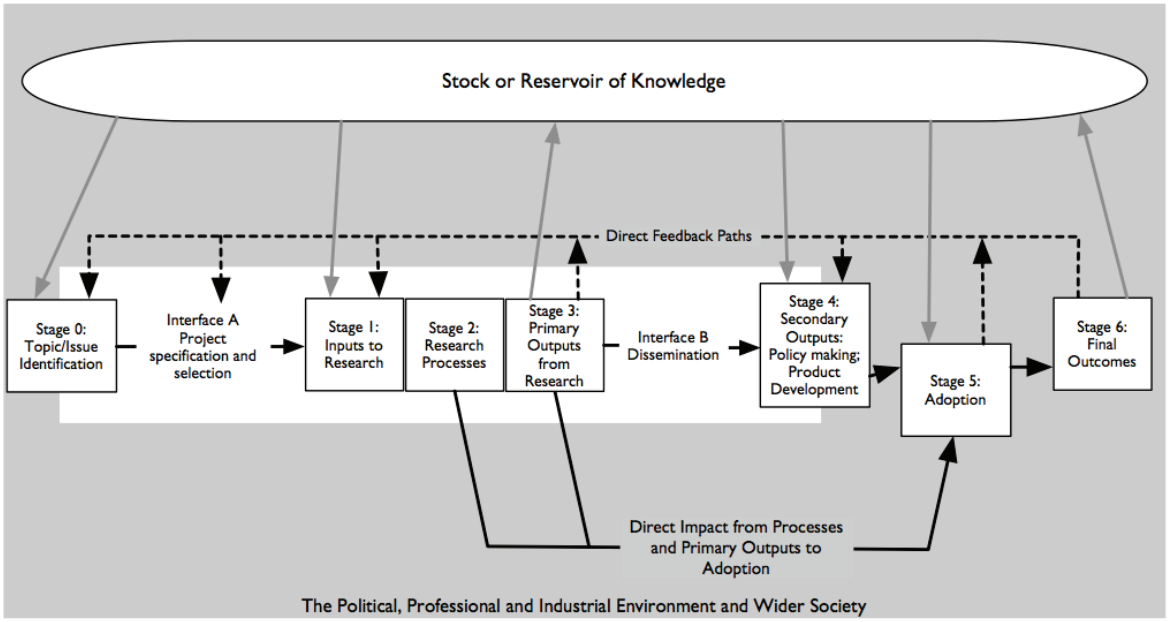


Figure 2-1 Logic model used in the Payback Framework. Adapted from Hanney et al., 2003.

The framework has five categories of impact: knowledge production, research targeting and capacity building, informing policy and product development, health and health sector benefit, and broader economic benefit (described in detail below), and these are used for the basis of our assessment of the level of impact of different case studies in the ranking exercise. They underpin our analysis of success factors.

- Knowledge production

This category covers the knowledge produced as a result of the research conducted, and this knowledge is in general captured in publications. Peer-reviewed articles are generally the most common measure, but editorials, meeting abstracts, reviews and patent filings are other examples of knowledge production. Citation analysis is one approach to understanding and measuring the output in this category.

- Research targeting and capacity building

This category captures benefits for future research created by the research conducted both in terms of the direction of research and research priorities, and the building of research capacity in terms of infrastructure, skills and staff development.

- Informing policy and product development

This category captures the impact of research on health policy (illustrated by such things as citation on clinical guidelines) and on product development as findings are taken up by the

private sector for commercialisation (possible measures are licensing intellectual property, contract research work and public–private joint ventures, along with new start-ups).

- **Health and health sector benefit**

This category covers health benefits and other benefits for the health sector (such as improved efficiency or cost savings) resulting from the findings of the research being put into practice. This typically occurs via the uptake of the policy, products or processes outlined in the previous category.

- **Broader economic benefit**

This final category covers the wider socioeconomic benefits resulting from the research. They might be the outcome of the increased productivity of a healthier workforce resulting from the health benefits described, or might result from increased employment or the development of new markets, stemming from the development of new products or processes. This can be very challenging to measure owing to the challenges of attributing such change to a particular piece of research.

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, knowledge production outputs are typically more common than outputs that inform policy. On the other hand, the outputs from a clinical research project may 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 be expected to require a longer period of study than a similar evaluation of clinical research. This is one reason for the extended timescale of 20 to 25 years in this study.

Typically, we term knowledge production and research targeting and capacity building 'academic' impacts, reflecting the fact that their impact is largely internal to the research system, and informs policy and product development, health and health sector benefits and broader economic benefits as 'wider' impacts, reflecting their wide impact on society at large. This conceptual dichotomy was supported by the empirical findings in the Cardiovascular Retrosight study described below, where it was observed that the highest correlations were between the first two, and the last three, categories of impact.

2.2 The Retrosight methodology

This study uses and develops the Retrosight methodology, which has been developed over the course of two studies conducted recently by RAND Europe. Both of these studies have focused on identifying factors associated with effective translation and downstream impact on health and society more widely.

The first of these RAND studies is Cardiovascular Retrosight, a multicountry study looking at the translation of cardiovascular research into practice. The methodology used there is comparable to the approach used in this study, although we have made some improvements to the methodology as well as adaptations to allow for differences between cardiovascular and mental health research. The approach used in Cardiovascular Retrosight is detailed in the methodology report for that project (Pollitt et al., 2011). One of the key

differences between this study and the Cardiovascular Retrosight project is the inclusion of the backward-tracing stream in the study. The aim of the backward-tracing stream of research was twofold: first to help to ensure that we considered some research that had been successfully translated, as there were concerns that improvements in mental health care resulting from research have been limited and hence it might be difficult to identify research that had been translated; and secondly to analyse the more downstream elements of research translation, which can be difficult to understand in detail when starting from specific bodies of research. Therefore, this change reflects both the difference in the research field and a refinement of the study methods.

Prior to the Cardiovascular Retrosight study, we conducted a similar study in arthritis research. That study used an earlier, less sophisticated, evolution of the Retrosight methodology based on stratified selection and a case study approach. All three studies share the use of case studies based on the Payback Framework.

2.3 Adapting the Retrosight approach for this project

As described in the following chapters, several changes and developments were made to the methodology to adapt it for use in this project. Several of these reflect pragmatic constraints and considerations. The study focuses on research in three countries: the USA, the UK and Canada. These were defined by the country of origin of the funders of the research. The work focuses primarily on schizophrenia, again representing the interests of one of our key funders. We also had some limitations in terms of the data available. For example, we used research clouds as the unit of analysis for our case studies in the forward-tracing stream of the research, as described and defined in the next chapter. This was partly due to their conceptual advantages as described there, but also due to difficulties in accessing grant information for one of the countries studied. Further constraints resulted from the availability of data in terms of our selection process, as described in that chapter.

Adapting the approach for mental health also raised some concerns, especially given the perception that changes in mental health practice have been limited. There was the possibility that it would be hard to identify further downstream impacts. Mental health research is also very diverse, including elements of psychological, sociological and environmental factors as well as more straightforward biomedical factors. For this reason, and to explore ideas we had about the development of research clouds, we conducted a pilot study to develop and test the approach. The findings of the pilot and the implications for the study are described elsewhere (Wooding et al., 2010). However, the key outcome of that pilot was the methodology described in this report.

CHAPTER 3 Case study selection

3.1 Unit of analysis: developing the concept of research clouds

The first step in selecting case studies was to determine our unit of analysis. In previous studies we have used research grants as the unit of analysis for payback case studies. Others have used a specific publication as the unit of analysis. Both of these approaches have some drawbacks, largely due to the fact that researchers do not think of their research in that way. Typically they pool various funding sources into one pot to support a range of research they are conducting, so attributing outcomes and impacts to a particular source of funding is challenging. Similarly, publications are often hard to consider in isolation, since there may be a series of publications on a particular cohort of patients, for example, or related to a particular trial or series of experiments. Attributing downstream impacts, or looking at inputs to any one individual publication, may be challenging. Therefore, in this study we attempted to use an approach which was more attuned to the way in which researchers think about their research and to the way in which that research is actually conducted.

Research clouds are a unit of analysis that embraces the idea that the process of research cannot be separated cleanly into discrete units. Unlike the units of papers or research grants, the idea behind research clouds is to reflect how researchers think about the division of their research while still providing a mechanism to isolate units of research for analysis. Research clouds represent units of research that share some sense of shared purpose or may in some way be delineated from the other work conducted by the researchers at the time.

Through the interviews in our previous work it became clear to us that researchers tend to see their work developing incrementally. They might identify a receptor responsible for physiological function and then move on to investigate its signalling, or learn and apply a new technique to a problem before moving on to apply that knowledge in a different experimental system (Wooding et al., 2004; Wooding et al., 2011). Both of these examples contain two successive research clouds in the story of the research on a particular topic. We chose the term ‘clouds’ to reflect the idea that each might have somewhat diffuse boundaries. In practice we found that although adjacent clouds might have some overlap, on most occasions we could define a workable delineation between them. This approach to defining the unit of analysis is more akin to approaches that define research ‘events’ or ‘breakthroughs’ – although we focused more on the stages in research activity than on the outputs/outcomes (IIT, 1968; Sherwin and Isenson, 1967). As shown in Figure 3-1, each

research cloud may be supported by a number of research grants and may produce more than one research paper.

Explicitly acknowledging the diffuse and contestable nature of the boundaries of the research clouds provided us with a framework to consider this in the case studies and ensured that we remained sensitive to the difficulty of defining boundaries in research

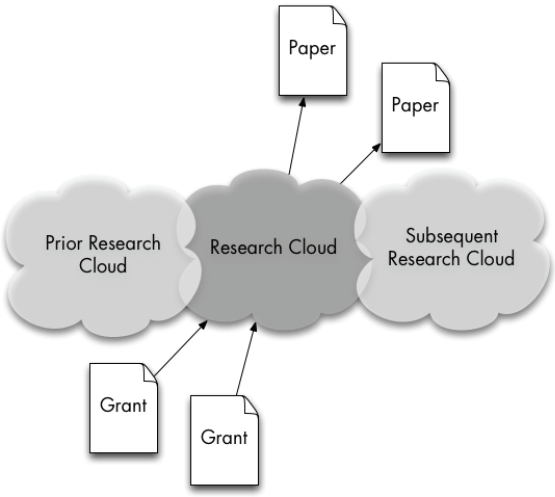


Figure 3-1 A diagrammatic representation of research clouds

As described below, we identified clouds through highly cited papers. However, the clouds were then delineated by talking to the corresponding author of the highly cited paper. It should be noted that these papers were just a route into the research cloud, which was more broadly defined in discussion with the researchers involved.

3.2 Aims of the case study selection process

The aims of the selection process were to generate a set of case studies covering a range of research from basic to clinical and interventional across our three study countries. This would allow us to analyse the factors that are important in the translation of research in mental health. This chapter sets out in detail the bibliometric techniques we used to minimise subjective bias in our case study selection, while ensuring that we produced a set of case studies that met our requirements, as laid out below.

Given the perception that mental health research has led to little progress, we were concerned that few of our case studies might show the downstream steps of translation into practice if selected randomly from across all publications at any given time. Therefore, we decided to focus on research that was considered a hot topic at the time it was conducted – that is, it included research which was highly cited in the years following publication. In addition, given the concerns over translation, we wanted to ensure that sufficient time had passed since the research was conducted for translation to have occurred. This timescale was likely to differ between basic and clinical research since basic research is typically further removed from its practical application, but is likely to be at least 12 years in the mental health research field, for the time between research being conducted and citation on clinical guidelines (Health Economics Research Group et al., 2008; Slote Morris et al.,

2011). We had to strike a balance between research which was as long ago as possible to allow for translation, but that was recent enough for researchers still to be available for interview and able to recall the research process. Therefore, we decided to focus case studies on research that was published during the time period 1985–90, and selected the research clouds by using bibliometrics to identify highly cited individual papers from this period. These papers were then used as a way in to the research cloud, and served as a starting point for discussion with the researchers in order to delineate the wider research cloud.

3.3 Bibliometric methodology

Bibliometrics employs quantitative analysis to measure patterns of scientific publication and citation, typically focusing on peer-reviewed journal papers (Ismail et al., 2009). In the current context we used bibliometrics to identify hot research topics for the forward-tracing case studies. In doing so, we made the working assumption that a highly cited research paper published in a peer-reviewed journal is a good indicator of an important piece of research.

We collaborated with Vincent Larivière at Observatoire des Sciences et des Technologies (OST) at the Université du Québec à Montréal (now at the Université de Montréal), an expert in bibliometric analysis. Using Thomson Reuters' Web of Science, we identified and classified a shortlist of highly cited papers in neuroscience and mental health research published between 1985 and 1990. The shortlisted papers were then examined and further filtered to finalise proposed case studies for the project.

The process of identifying case studies broadly consisted of six stages, as follows. These steps are described in more detail in Appendix A.

1. Identify a set of journals from which to search for papers. All neuroscience and mental health research papers with a US, UK or Canadian address published between 1985 and 1990 identified from the Web of Science were included.
2. Group the papers by research type into 'basic' and 'other', where the 'other' group refers to more applied research. The approach used was based on the National Science Foundation's 'Research level' classification system (Narin et al., 1976).

It should be noted that although we selected a range of pieces of research across these research types, this categorisation was only applied to the initial paper selected for each case study. In some instances this may mean that the research cloud as a whole does not fit into the same category, and we addressed this by reclassifying the final case studies before analysis. However, the key aim of this approach was to ensure that we selected a suitable range of research and not to create a rigid system of classifying case studies.

3. Identify highly cited papers, defined as the papers that fell between the 85th and 95th percentiles in terms of citations for their groups ('basic' or 'other').

We chose this range for the following reasons:

- We wanted to exclude outliers, and so not select the most highly cited. This was due to the sense that the very highly cited papers (above the 95th percentile)

typically have something unusual about them and are not representative – and hence are less able to provide lessons that may be applied widely.
- We needed to ensure that we were comparing like with like in our paper selection and thus wanted a relatively tight percentile range.
- We needed to ensure that we could select at least two papers from Alberta, as AIHS was supporting two case studies.

4. Group the papers by country, using the address information of the first author of each paper.
5. Filter papers to check that they are in scope, separate the 'other' group into clinical and interventional, and further subdivide the interventional group into biological, psychosocial and health services intervention. This consisted of a combination of searching for the key term 'schiz' in the title, and carrying out a manual review of titles.
6. Populate a matrix with the remaining papers using the country and research-type classifications described above. Then select papers within each group, randomly where possible, and pragmatically where necessary.

Pragmatic selection was necessary in some cases – for example where we were not able to contact the first or corresponding author on a particular paper.

The final selection and distribution of the 18 papers across country and classification is shown in Figure 3-2. In this matrix we include the title of the paper and the number of citations in the five years following publication.

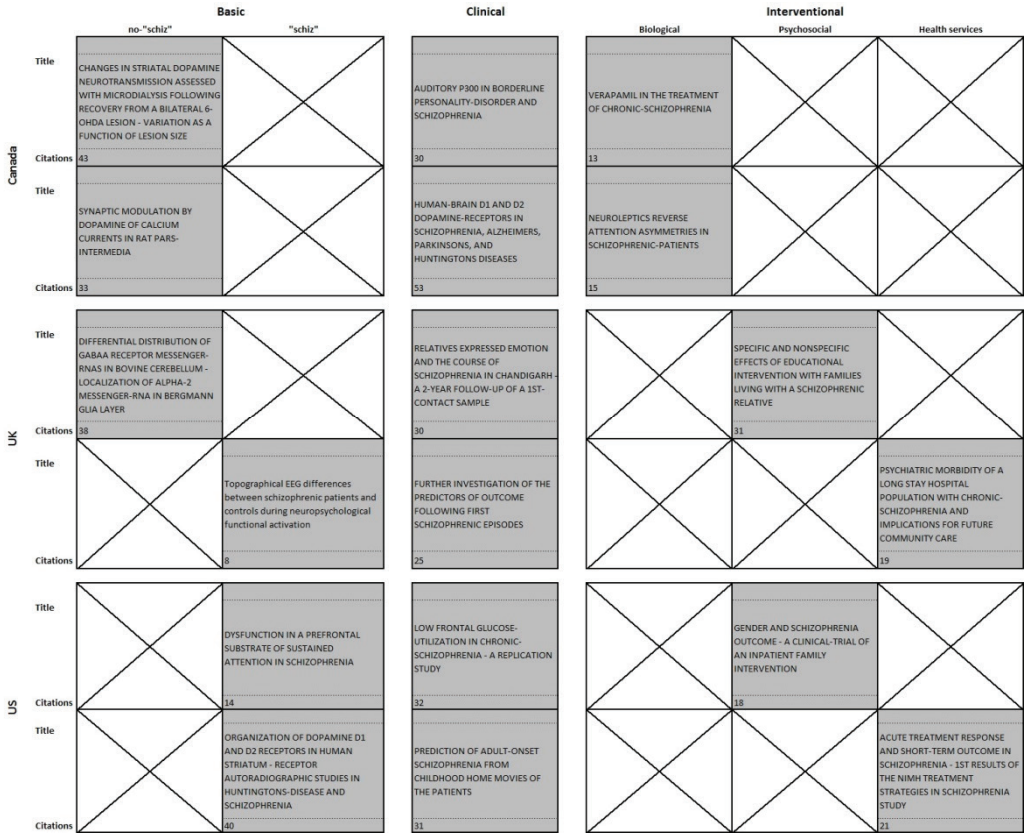


Figure 3-2 Case study selection by country and research type

Having selected our full set of 18 case studies, we were able to contact the researchers involved in order to set up interviews and begin the process of developing the case studies.

3.4 Strengths and weaknesses of the approach

One of the key strengths of this approach is that it ensures that all published research is available for inclusion, and does not restrict the types of research that will be identified. For example, in the Cardiovascular Retrosight project, we identified case studies using a list of grants from research funders. This was a useful and straightforward way to identify research projects, but it limited the scope of the study to research that was supported by direct grant funding by a public health research funder (and in fact those funders who provided findings records). This approach, using publications to access research clouds, allowed access to, for example, industry-supported research, or research which was conducted using discretionary funding. As such, this was a more complete set of the research which was conducted and was limited only by that research which led to publication in the academic journals indexed.

Another key strength was that the approach was data driven, with a selection process designed to avoid any bias or influence on the part of the research team, preventing cherry-picking of interesting research or well-known researchers for the case studies.

One of the weaknesses of the approach was the lack of accuracy of classification of the papers and the resulting clouds into the various categories we had identified for

stratification of our case studies. This may result both from inaccurate classification of the original paper (e.g. if the content of the paper is more clinical or basic than understood from the abstract review used to classify them into categories), but more commonly will result from the research cloud not fitting the same classification as the publication used to identify it. For example, though the first author of the publication may list a Canadian address, it may be that the majority of the research in the cloud, and much of the other research in the cloud, is based in a different country. The choice of definition of clinical and basic research also changed slightly between the initial selection of papers and the final classification of case studies, which has changed the number of case studies falling into each group. However, overall the strategy was successful in generating a mix of research across research levels and countries, as desired.

	Canada		UK		USA	
	Refused	Unable to contact	Refused	Unable to contact	Refused	Unable to contact
Basic	0	0	2	4	1	0
Clinical	1	3	0	0	0	0
Interventional	0	0	0	0	1	2

Figure 3-3 Case study refusals, and instances where we were unable to contact the researcher

A second limitation of the approach was that in several cases we were unable to carry out a case study on a selected publication because we were unable to contact the researcher identified as the first author of the publication. These are shown in Figure 3-3. This may have been for several reasons: because the researcher had retired, they had died, or they had moved out of the field and contact details were no longer available. The rates of these occurrences shown in the table suggest that the 20 to 25-year time span since publication used is approaching the maximum time lag that can be explored with this type of case study methodology. Also, our sample may be skewed by refusals of researchers to participate in the project, although the number of refusals we received was small.

To some extent, this pragmatism in the final stages of the selection process seems at odds with the stringency of some stages of the selection process. However, this was driven by a desire to avoid cherry-picking – as described above – while still ensuring that the final selection of case studies included those with the potential for downstream impacts while responding to the need to select a sample of case studies that met the study requirements in terms of coverage of types of research and countries of origin of the research. Therefore, although this approach appears stringent in terms of its criteria, it was also pragmatic in ensuring that our sample was suitably representative.

CHAPTER 4 Perspectives selection

4.1 Overview of selection

For the perspectives, we wanted to identify recent advances that had resulted in substantial benefits to patients with schizophrenia, or that showed clear potential to change care. This would allow us to look back and hopefully find examples of where research *has* been successfully translated to treatments that have benefited patients. We developed six perspectives for this project.

In order to investigate a suitable range of advances in care we developed a set of criteria to aid our selection. In summary, we wanted to consider the following:

- A balance of pharmacological and non-pharmacological interventions.
- Some advances that focused on the individual patient and others that focused more on their environment (family, community, etc.).
- Advances that were at different stages of translation – from the cusp of entering practice to treatments which had reached mainstream treatment.
- Some treatments that had been similarly adopted across the project countries and some that had not (i.e. where use was more widespread in one or two countries than in others).

Our initial approach to selection was to conduct a Delphi-like survey of various mental health stakeholders to generate a broad list of possible topics for perspectives. Because of the low response rate to the Delphi survey, it was supplemented with a review of clinical guidelines and the opinions of the project's expert advisors.

When we reviewed the results of this survey, two immediate candidate advances emerged for patients with schizophrenia: cognitive behavioural therapy (CBT) and 'early intervention'.[3]

This led to the clear selection of the first two perspectives:

[3] This is difficult to define straightforwardly, but broadly it refers to the provision of treatment either immediately after the first onset of psychosis, or before the first onset of psychosis when initial warning signs are detected.

- **CBT:** This emerged as a clear priority in Canada and the UK, and was the most commonly suggested non-pharmacological intervention.
- **Early intervention:** This was again a very popular suggestion, particularly in Canada and the UK.

The grouping of other responses was less clear cut and so we reviewed clinical guidelines for schizophrenia treatment in the three project countries. This review aimed to provide a more detailed set of pharmacological advances to consider – in many cases, survey respondents had simply suggested 'antipsychotic medication' or something similar. The review produced a second list of interventions which we mapped against the original list from the survey. Looking at the country of each survey respondent, we then ranked the guideline list for each country to determine which suggested advances had gained the most support in each.

In addition to the two perspectives described above, the survey responses suggested two areas in which to select perspectives: interventions with more of a community or family focus (one perspective) and the broad area of pharmacological interventions (three perspectives, to ensure a balance of different kinds of treatment). We took the suggestions relevant to these two areas from the Delphi exercise and guideline review and discussed them with the project's expert advisors: Professor Shitij Kapur and Professor Harold Pincus. At this point our criteria for selection were that the intervention needed to be conceptually well defined across the three project countries (to ensure that we compared equivalent treatments), feasible in terms of scale and likely accessibility of information, and that, if possible, we should look at interventions at different stages of uptake (i.e. a consideration of how recently they had been developed).

This consultation resulted in the selection of the following:

- **Supported employment:** This was deemed to be a tractable and interesting story with a substantial evidence base for the effectiveness of the intervention. It also provided some balance in the non-pharmacological perspectives, in that initial development appeared to have taken place primarily in the USA.
- **Clozapine:** The antipsychotic clozapine was a fairly common response in the Delphi exercise, and although its origins lay further back in time than we had initially planned to look, it was felt that more recent developments in its use warranted its inclusion as a perspective.
- **Addressing metabolic side effects of antipsychotics:** The rise in the use of atypical antipsychotics has brought a greater awareness of the metabolic side effects that these drugs sometimes have. This perspective looked at changes in pharmacological practice as a result of that.
- **Addressing cognitive deficits in schizophrenia:** There has been growing recognition in recent years of the cognitive deficits associated with schizophrenia, either as a core part of the disorder or as a result of antipsychotic medication. This

perspective focused on identifying and measuring these deficits, and on attempts to develop pharmacological treatments to address them.

The above set of perspectives ensured that we covered various aspects of pharmacological treatment (three perspectives), psychosocial interventions (CBT), interventions focused on living in the community (supported employment) and advances in the way in which mental health services are delivered (early intervention). Within the pharmacological perspectives we covered a substantial timeframe, with much of the development of clozapine occurring in the 1960s and 1970s, attempts to address metabolic side effects coming after the introduction of atypical antipsychotics in the 1990s, and a focus on cognitive deficits much more recently.

Details of the Delphi-like survey and the review of clinical guidelines are provided in the following sections.

4.2 Delphi-like survey

The aim of the survey was to identify key treatments, advances or interventions that could legitimately be said to have improved the care and management of schizophrenia and which would also lend themselves to further research and analysis as perspectives. Although the national treatment guidelines for schizophrenia in Canada, the UK and the USA contain information regarding current recommended treatments and interventions for schizophrenia, we wanted to select advances and treatments that had already brought significant benefit to the lives of patients. Given the complexity of mental health treatment and the lack of major advances in care, we decided that a Delphi-like survey would yield an accurate and revealing picture of the treatments that have had a genuine and significant impact on the lives of patients with schizophrenia.

A Delphi survey may be an effective means of gathering qualitative information, in this case the opinions of stakeholders on the key advances or treatments in schizophrenia care, in a structured way. In its most typical form, a Delphi survey involves using a questionnaire that asks participants to list, rank and rate a series of items over a number of rounds interspersed with feedback collection. The aim in most instances is to drive participants to consensus on a set of issues, factors or events, but the method may be used in a more open-ended manner to reveal a range of options instead (Ismail, 2009).

We attempted to use the Delphi methodology to gauge opinion from mental health researchers, practitioners, service-users, nurses and other stakeholders on what they considered to be the most significant advances and treatments in the care and management of schizophrenia. To do this we needed to pose a question which would prompt respondents to consider all elements of a schizophrenia patient's care today and identify the elements of that care that have genuinely brought benefit and positive outcomes to the patient. The difficulty was choosing language which would be meaningful across Canada, the UK and the USA and across different professions and levels of engagement with the issue of schizophrenia care. For example, although we were initially interested in identifying breakthroughs in schizophrenia care, we were concerned that by using that terminology we would preclude respondents from suggesting other elements of

schizophrenia care that could not be considered to be a breakthrough but which have, nonetheless, yielded considerable benefit to patients. As a result, we consulted our expert advisors for guidance on suitable wording and phrasing for the survey question and finally settled on the following:

'What are the most important interventions that have been introduced into practice for the treatment of schizophrenia in the last 5–10 years and yielded significant benefits to patients?'

A list of potential respondents to the survey was compiled from the following sources:

- steering committee members and their contacts;
- the research team;
- internet searches for key organisations and individuals with an interest in mental health and schizophrenia.

In total, the Delphi-like survey was sent to 113 people. Table 4-1 shows a breakdown of invited respondents by country and the rate of response received.

Table 4-1 Rate of response by country

Country	No. sent survey	No. respondents
Canada	28	9
United Kingdom	34	14
United States	39	7
Other	2	1
Unknown	10	4

We invited individuals from a wide range of schizophrenia and mental-health-related organisations across the areas of research, treatment, support and advocacy to respond. Individuals represented researchers, clinicians, service-users, service-providers and nurses. From the outset, we aimed to get as broad and varied a group of respondents as possible and we were particularly keen to ensure that service-users were represented in the sample. The survey was administered electronically and all prospective participants were invited to forward the survey to any additional contacts they felt would be interested in taking part in the study. Assuming all of the responses were from people we directly approached, our overall response rate was 31%. Figure 4-1 shows the breakdown of respondents by profession.

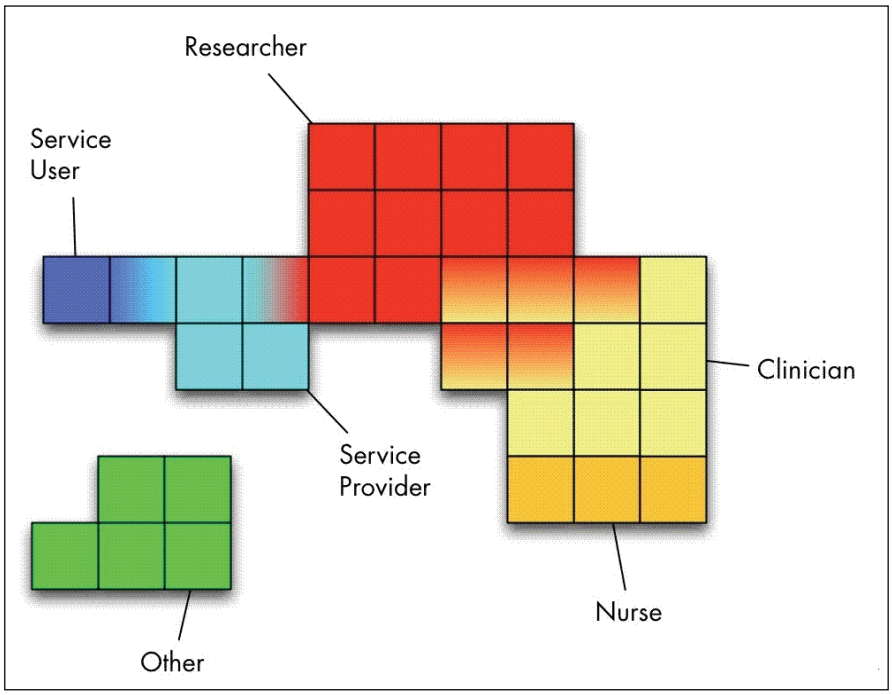


Figure 4-1 Breakdown of respondents by occupation

As Figure 4-1 indicates, the majority of our respondents were either researchers or clinicians (or both). Despite inviting a number of support service organisations and patient groups to participate in the survey, we were unable to get greater representation from service-users and support-service-providers.

4.3 Review of national clinical guidelines

To supplement the Delphi responses and on the advice of steering committee members, we undertook a comparison of the treatments and interventions discussed in the national guidelines for schizophrenia in Canada, the UK and the USA. The guidelines consulted were the American Psychiatric Association Practice Guideline (USA), the Schizophrenia Patient Outcomes Research Team (PORT) Treatment Recommendations (USA), the National Institute of Clinical Excellence clinical guidelines (UK) and the Canadian Psychiatric Association Clinical Practice Guidelines (Canada).

Comparing the clinical guidelines with the responses from the Delphi-like survey resulted in the final list of potential interventions shown in Table 4-2 below.

Table 4-2 List of interventions/treatments for the second round of the Delphi-like survey

1	Family support interventions
2	Psychoeducational interventions
3	Vocational rehabilitation, including supported employment
4	Community treatment systems, including Assertive Community Treatment

5	Alcohol and substance abuse programmes
6	Peer support and peer-delivered services
7	Cognitive remediation
8	Psychosocial interventions for weight management
9	Early intervention
10	A culture of hope and recovery, including the recovery model of care
11	Cognitive behavioural therapy
12	Reduction of stigma – i.e. greater public understanding and tolerance of schizophrenia
13	The use of antidepressants for depressive symptoms
14	Avoidance of antipsychotic polypharmacy
15	Management of side effects with first-generation antipsychotics by reducing dosage
16	Management of side effects with first-generation antipsychotics by switching to a second-generation agent
17	Use of clozapine in treatment-resistant patients
18[4]	Use of second-generation antipsychotics as a frontline treatment
19	Management of metabolic/cardiovascular side effects of (second-generation) antipsychotic agents

4.4 Combining the Delphi-like survey and guideline review

The two sources described above were combined in making the final perspectives selections. The responses from the Delphi-like exercise were mapped as closely as possible against the interventions recommended in the practice guidelines. The Delphi survey gave us an idea of the areas of importance in a general sense, while the guidelines allowed us to narrow that down to a specific intervention which could be investigated in detail, so the two approaches were complementary. The resulting list was then discussed with our expert

[4] These two interventions were included at the suggestion of our expert advisors, who considered them to be major advances that had not been captured specifically from the clinical guideline review or Delphi-like survey. We felt it would be wise to work with this wider list initially to ensure that no major advances were excluded, and then narrow it down during the selection process.

advisors in the context of the guiding criteria we had developed (a balance of pharmacological and non-pharmacological interventions, variations in uptake in project countries and differing focuses of the interventions). We also considered the feasibility of constructing a comprehensive perspective on each intervention, in terms of the scale of the topic, how well defined it was across the project countries, and whether there would be sufficient information available to develop an interesting and useful narrative.

This gave us our final set of six interventions and, as with the case studies, we began a scoping literature review and contacted relevant interviewees.

4.5 Strengths and weaknesses of the approach

The final approach we used has a number of weaknesses. The first of these are the low response rates and the relative response rates in different groups in the Delphi study. There was a reasonable spread of respondents across countries, although the response in the UK was somewhat higher than in the other countries. This is notable since treatment is likely to be fairly consistent across the UK, while there may be more variation from region to region, for example, in Canada – as such it would have been valuable to get a wider range of viewpoints in Canada. However, more notable was the low level of response from service-users and patient representative groups. The majority of responses were from clinicians and researchers.

The process of identifying suitable subjects for perspectives made clear to us the diversity of opinion in the mental health field, and it is likely that there is no agreed list of six advances that would be generally accepted. Our approach has tried to draw on a variety of sources to select a varied set of advances that should mirror the variety of mental health treatment options. We have ensured input from a range of stakeholders and across the three countries, and worked hard to ensure that we phrased our questions to allow a wide range of interventions to be considered.

CHAPTER 5 Building the case studies

5.1 Structure of case studies and the Payback Framework

This chapter describes how we developed a detailed case study round each research cloud. We start by outlining the framework to be used for the case studies. As described in Chapter 2, these case studies are based around the Payback Framework, which has been used previously for many such studies – including the Cardiovascular Retrosight study – and it offers an effective approach to developing case studies. The impact categories form the basis of much of the analysis conducted, and have already been defined in detail in Chapter 2. As an overview, the five categories are as follows:

Knowledge production: Covers the knowledge produced as a result of the research conducted and primarily encompasses a range of publications.

Research targeting and capacity building: Captures benefits for future research created by the research conducted.

Informing policy and product development: Captures the impact of research on health policy and on product development as findings are taken up by the private sector for commercialisation

Health and health sector benefits: Covers health benefits and other benefits for the health sector resulting from the findings of the research being taken up and put into practice.

Broader economic benefits: Covers the wider socioeconomic benefits resulting from the research.

The logic model forms the basis of the case study structure. The way in which it has been applied in this study is described in detail below.

5.1.1 The logic model and case study structure

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 5-1). 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.

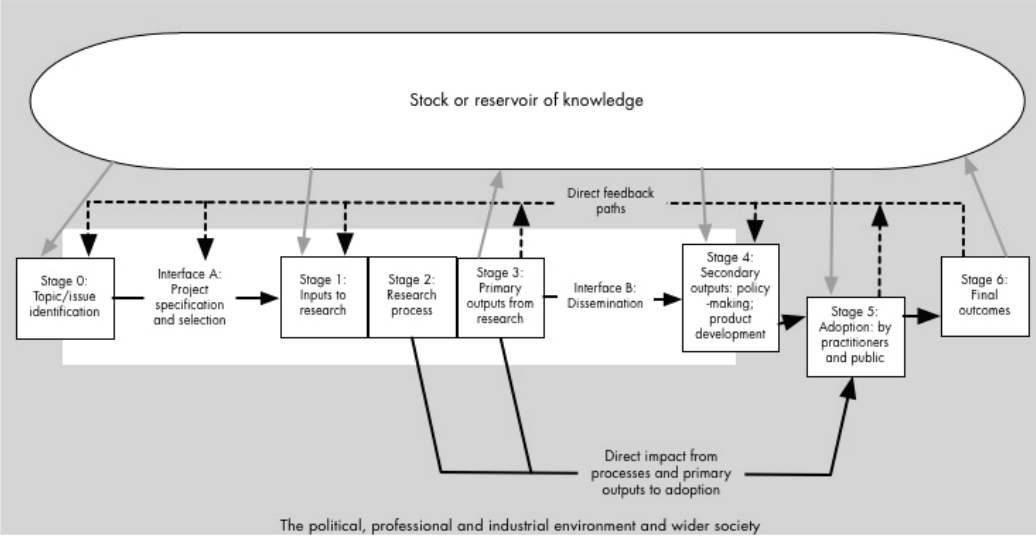


Figure 5-1 The Payback Framework's logic model of the research-funding process. Source: Hanney et al., 2003

There are seven stages in the logic model, from the initial idea for the research project through to the wider impact on society, as shown in Figure 5-1. We used these stages to help define a structure for our case study narratives. The aim of this approach was to produce a set of case studies which were comparable and contained equivalent breadth and depth of information, despite being produced by different authors and covering potentially significantly different material. This was important as it allowed comparison between the case studies, both for the effective assessment of the level of impact of each research cloud at the scoring stage, and of the factors which might contribute to translation when coding the case studies. Based on the Payback Framework, we developed a structure for the case studies consisting of 13 sections. This is a larger number than the seven stages (plus two interfaces) defined in the framework, and most of the additions were to provide contextual information or to account for additional information needed because of the methodology used in our study. The sections in the case studies are described below, with reference to the relevant section of the Payback Framework where appropriate.

5.1.2 Case study sections

Summary

The first section of the case studies is a brief summary. This is a short section which acts as an abstract for the rest of the case study and helps to orient the reader. This was added to reflect the multiple ways in which the case studies was used. In particular, this was helpful for scorers to identify the case studies in the scoring process.

Introduction

This is split into three subsections: scientific background, researchers' background and institution background. The section was intended to give some background to the case study and to set it in context. This served several purposes. First, it allowed for the scientific background needed to understand the content of the case study to be explained to the audience, which in this case was non-expert. It also gave a centralised place to set out some of the contextual information such as the career stage of the researchers involved and the location in which the research which is conducted – which was useful for the analysis

stage. Finally, it provided a 'glossary' upfront of the names of the key researchers involved in the work for reference which was useful for the reader to refer to when these researchers were referenced later in the case study.

Defining the research cloud

This section was introduced to reflect the unit of analysis used in this study. It is a short section which lays out the content of the research cloud. As this is a flexible concept that allows us to define a body of research in a similar way to how the researchers themselves think about their work, it is necessarily differently defined by different researchers in different fields and as such needed to be set out for each case study. This section typically consisted of a description of what is in and out of scope in terms of the cloud, any complications in this and, in many cases, a list of the research outputs that defined the knowledge output of the cloud. These are not described in detail here, however – that is kept for Stage 3.

Stage 0: Opportunity Identification / Research Needs Assessment.

This section corresponds closely with the Stage 0 as defined in the Payback Framework, Topic/issue identification, and was used to describe how the idea for the research was born. There are numerous potential sources for 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 (e.g. 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.

The section was divided into three subsections: inspiration, feasibility and potential value.

Stage 1: Inputs into research

This stage concerns the resources available to the project and is broadly defined as in the Payback Framework. Possible resources considered include financial support, human resources (such as scientific expertise and administrative support, and also the knowledge and skills of the research team), physical resources (scientific equipment, facilities and general consumables as well as samples/patients) and collaborators.

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 may be helpful to explore the extent to which collaborators and potential users are involved in implementing the research as they often influence how the research evolves.

Stage 3: Primary outputs

The primary outputs from research fall into the first two payback categories: knowledge production, and research targeting and capacity building, and these two subsections are used in the case study structure.

Most knowledge production outputs (which will largely be publications, but may include other types of output) 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 policymaking 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. As such, the second part of this section, entitled 'Targeting future research', was split into two further subsections: effect of the researchers' careers and future work, where future work could be within the same team or group, or more widely in the field or other fields.

Interface B: Dissemination

Dissemination is more than 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 may 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 and interactions with the media.

This section was subdivided into two categories: academic dissemination and wider dissemination, reflecting the differences in focus and in outcome between these types of dissemination.

Stage 4: Secondary outputs

Secondary outputs contribute mainly to the third payback category: informing policy and product development. Unlike primary outputs, secondary outputs may 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 may be found in the form of research citations in documents produced by professional and public policymaking bodies, such as clinical guidelines. Research findings may 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 may influence academic course curricula in medical schools and may 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 may be attributed to the specific research findings, as opposed to other factors (such as a more general change in the 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 may accrue over a protracted period of time and are the most difficult to attribute to an individual research cloud.

Table of payback

We included a section in each case study which summarised the key impacts identified across the five impact categories. This served as a useful summary of the case study and

was particularly helpful for the scoring process.

Timeline

Finally, we included in each case study a timeline which set out the major developments and steps in each study so that the narrative could be visualised. This helped to ensure clarity in terms of the order in which events occurred and the possible causality between them.

5.2 Data collection

Data collection consisted of three main elements:

- Interviews, primarily with researchers involved in the work or working in the field at the time the research was conducted.
- Desk-based research, reviewing the academic and wider literature and other relevant documentation.

- Bibliometric analysis of the publications arising from each research cloud and the networks of the researchers involved in the work.

Each of these is described in more detail below.

5.2.1 Interviews

The most important element in building the case studies was conducting detailed interviews with the researchers involved, collaborators and others in the field at the time. In total we interviewed 48 people across the 18 case studies, and several others contributed or commented by email. All initial interviews with the first/corresponding author were conducted in person by two RAND researchers (in a few cases one researcher joined in by telephone), while the same process occurred for follow-up interviews wherever practical. In instances where distance or scheduling constraints prevented us from conducting follow-up interviews in person, these were done by telephone. Interviews were conducted in a semi-structured format using a protocol based around the Payback Framework, which allowed us to ensure effective gathering of similar levels of information for all case studies while also giving researchers the flexibility to pursue interesting lines of enquiry, address gaps in knowledge or verify information collected in previous interviews and from other sources. The interview protocol used is available in Appendix B.

In all cases, the first interview conducted was with the corresponding author for the paper used to identify the case study. This was important as at this interview the scope of the research cloud was identified and agreed with the researcher – through discussion prompted by a list of publications for that author over the relevant period, and a network analysis of their collaborators at the time. This was a crucial first step as once the cloud was defined clearly, other key members of the research team involved – as well as wider collaborators, competitors and downstream users of the research – could be identified for further interviews. Often the cloud was broadly agreed at the interview stage, and then finalised through email discussion with the researchers involved after further interviews had been conducted, since there might have been relevant publications to be included in the cloud that did not include the initial researcher interviewed as an author. However, what was important was to define a conceptual understanding of what defined the cloud from the outset, so that the interview and wider research could be conducted with this in mind. As shown in Figure 5-2, there were typically around 5–15 publications associated with each cloud.

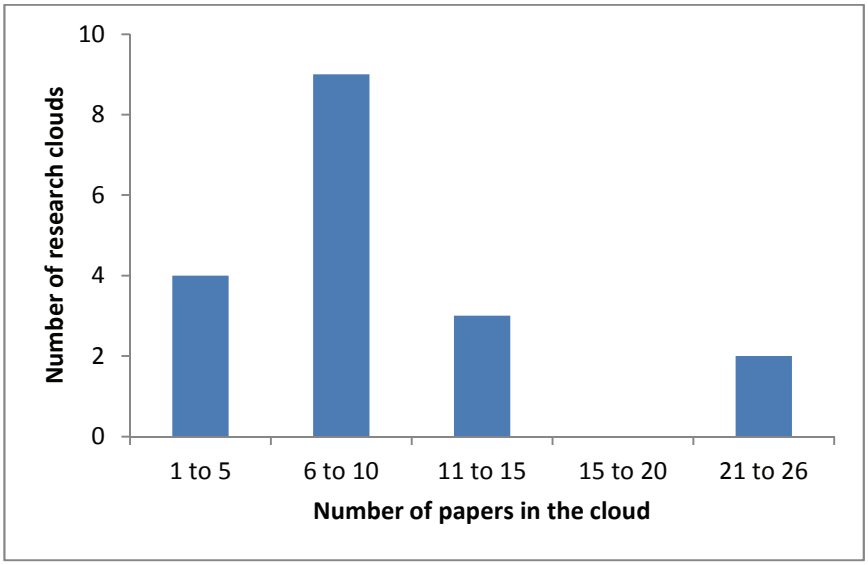


Figure 5-2 Histogram showing the number of papers in the research clouds

Before interviews, informed content was sought and all interviews were conducted in confidence. No direct quotes were used without the express permission of the interviewees. All interviews were recorded, with the permission of the interviewee, and were transcribed.

5.2.2 Desk research

Interviews were complemented by desk research. This consisted of a number of elements. A key component for all the case studies was a review of the scientific literature – both that coming out of the research conducted in the cloud and other relevant papers, reviews and textbooks which laid out the state of the field at the time – and the research of other research groups working in the area, so that the unique contribution of the cloud could be understood. The review also covered work which was influenced by the research cloud – for example, the subsequent work of the research team – and publications which cited those papers included within the research cloud, and sometimes further generations of publications subsequent to that. This was important for identifying the research targeting which resulted from the work, and also in tracing small but potentially material contributions to further downstream impact.

In addition, a review of other, wider documentation was conducted to identify wider impacts. This included, as relevant, review papers, clinical guidelines, policy documents, patents and even legal proceedings relating to outputs of the research. Sometimes this review led to further interviews with those involved in the uptake and dissemination of the work, or those who had used or implemented the research.

5.2.3 Bibliometrics

Each case study included a bibliometric analysis of the publications arising from each research cloud. This is one way of assessing the production and dissemination of knowledge from the research cloud. An example panel is provided in Figure 5-3. This analysis included the indicators defined in Box 1 below.

WoS indexed papers – The number of publications arising directly from the research cloud and included in the analysis (i.e. indexed in Web of Science)

Total citations of papers – Total number of citations received by the above papers

Percentage self-citation – Percentage of total citations (to journal articles) that are made by papers on which the first author is one of the of the authors of the cited paper

Mean citations per paper – 'Total citations of papers' divided by 'WoS indexed papers'

Cloud citations – Number of distinct citations received by the research cloud (i.e. removing duplication where a single article cites two or more papers from the cloud)

Second-generation cloud citations – Number of citations received by articles citing papers contained in a research cloud

Third-generation cloud citations – Number of citations received by articles citing papers that have cited papers contained in a research cloud

Citations in RCT / meta-analyses / practice guidelines / review articles – Total citations received by 'WoS indexed papers' in these publication types indexed in the WoS

Research level of cloud papers – Classification by journal of publication into four levels (as described in Hamilton, 2003): 'basic' (RL=4), 'clinical investigation' (RL=3), 'clinical mix' (RL=2) and 'clinical observation' (RL=1)

Publications per country – Distribution of the cloud's 'WoS indexed papers' by first address, typically the address of the first author

Specialty – The subfield (determined by journal) of papers citing the research cloud (i.e. of 'cloud citations')

Citations by country – Distribution by country (of first address) of 'total citations of papers' (including self-citation)

Citations by publication year – Number of citations (including self-citation) per paper, ordered by year of publication

Cloud citations by year – Number of citations (including self-citation) received by papers in the research cloud by year

Box 1 Definitions of bibliometric indicators

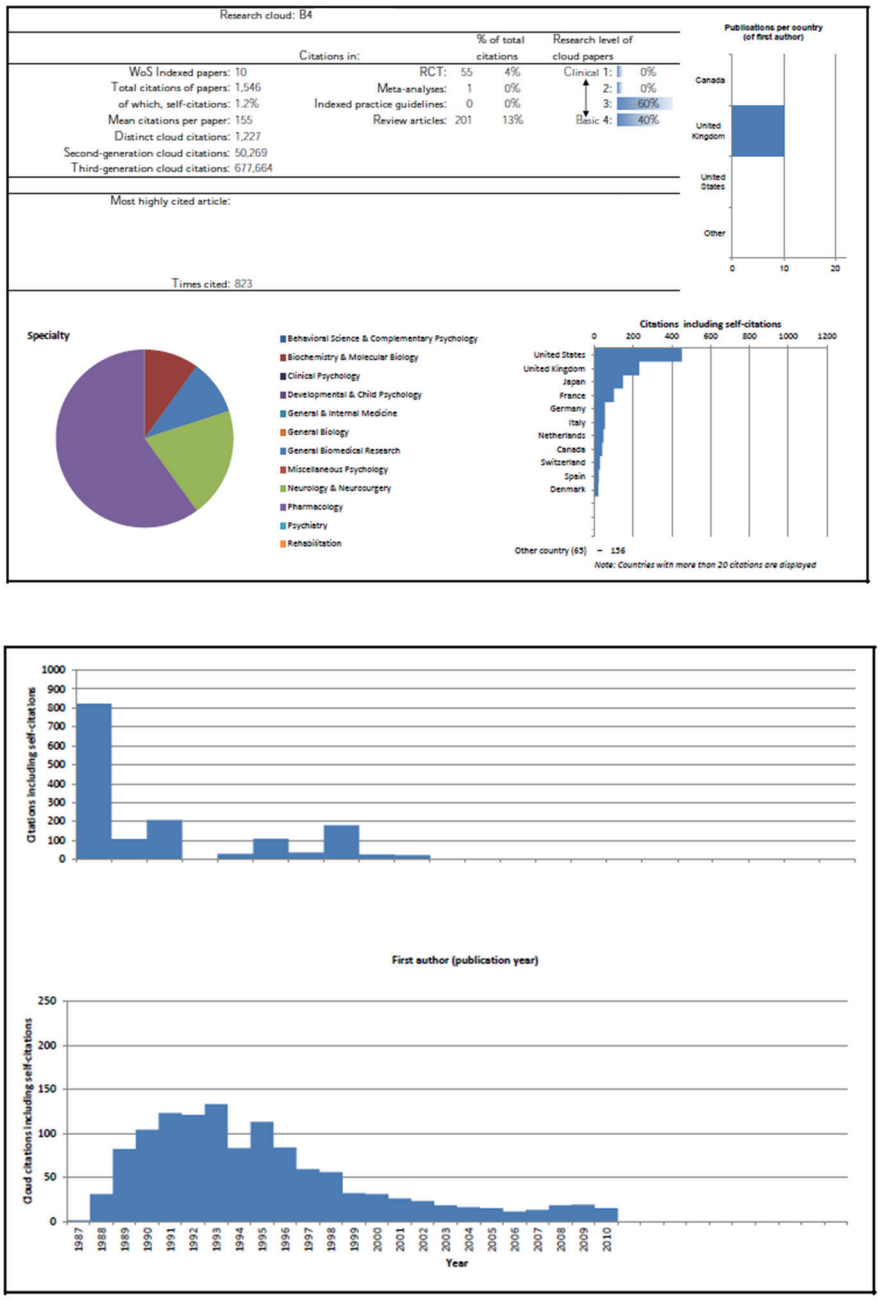


Figure 5-3 Example of bibliometric analysis for forward-tracing case studies

This analysis was useful not only in understanding the knowledge produced by the research cloud, but in identifying likely sources for further desk-based research to identify wider impacts of the research.

5.3 Validation

Draft case study narratives were shared with and cleared by all the researchers interviewed as part of the data-collection process. In addition, the narratives were independently peer

reviewed by experts in the field. The aim of this review process was to assess the accuracy of both the science described and the attribution of impacts to the research.

We aimed to identify two reviewers with relevant expertise for each case study, one from the same country as the case study, and one from another project country to ensure the wider implications of the research outside its country of origin were taken into account. In all we identified and approached 168 reviewers and received 27 reviews. The distribution of these reviewers across the case studies is shown in Figure 5-4. The assessments of the reviewers are summarised in Figure 5-5. Revisions were made to the case studies based on the reviewer comments as appropriate. Finally, the case studies were reviewed by one of the quality assurance reviewers for the project to ensure that the research team had addressed the comments of the external reviewers suitably.

Case study	Number of reviewers
1	1
2	1
3	2
4	1
5	2
6	1
7	2
8	2
9	2
10	1
11	2
12	1
13	1
14	1
15	1
16	2
17	3
18	1

Figure 5-4 Number of reviewers per case study

	Reviewer	Accuracy of the historical account	Accuracy of the statement of the impacts	Acknowledgement of the work of others	Description of the science	Are there inaccuracies which may affect judgement of impact?
1	Reviewer 1	Accurate	Slightly inaccurate	Minor omissions	Suitable	No
2	Reviewer 1	Slightly inaccurate	Accurate	No rating (Accurate)	Suitable	No

3	Reviewer 1	Accurate	Accurate	Accurate	Suitable	No
	Reviewer 2	Accurate	Slightly understated	Major omissions	Slightly inaccurate	Possibly significant
4	Reviewer 1	Accurate	Accurate	Accurate	Suitable	No
5	Reviewer 1	Accurate	Slightly overstated	Minor omissions	Slightly inaccurate	No
	Reviewer 2	Accurate	Accurate	Major omissions	Suitable	Hard to judge
6	Reviewer 1	Accurate	Slightly inaccurate	Accurate	Suitable	No
7	Reviewer 1	Accurate	Slightly overstated	Accurate	Suitable	Slight
	Reviewer 2	Accurate	Accurate	Accurate	Suitable	No
8	Reviewer 1	Accurate	Slightly understated	Accurate	Suitable	No
	Reviewer 2	Slightly inaccurate	Slightly understated	Minor omissions	Suitable	Slight
9	Reviewer 1	Accurate	Seriously understated	Minor omissions	Slightly inaccurate	Possibly significant
	Reviewer 2	Slightly inaccurate	Accurate	Minor omissions	Suitable	No
10	Reviewer 1	Seriously inaccurate	No rating (Accurate)	No rating (Accurate)	No rating (Suitable)	No
11	Reviewer 1	Accurate	Seriously understated	Accurate	Suitable	Hard to judge
	Reviewer 2	Accurate	Accurate	Accurate	Suitable	No
12	Reviewer 1	Accurate	Accurate	Accurate	Suitable	No
13	Reviewer 1	Accurate	Slightly understated	Accurate	Suitable	No
14	Reviewer 1	Accurate	Accurate	Accurate	Suitable	No
15	Reviewer 1	Accurate	Accurate	Accurate	Suitable	No
16	Reviewer 1	Accurate	Accurate	Minor omissions	Suitable	No
	Reviewer 2	Accurate	Seriously understated	Major omissions	Suitable	Possibly significant
17	Reviewer 1	Seriously inaccurate	Seriously understated	Accurate	Slightly inaccurate	Possibly significant
	Reviewer 2	Slightly inaccurate	Slightly overstated	Major omissions	Significantly inaccurate	Hard to judge
	Reviewer 3	Accurate	Accurate	Accurate	Suitable	No
18	Reviewer 1	Accurate	Not clearly stated	Major omissions	Suitable	Hard to judge

Figure 5-5 Summary of reviewer comments on case studies

5.4 Strengths and weaknesses of the approach

This approach had a number of strengths. The case study methodology allowed for a high level of detail to be gathered in relation to each research cloud while still retaining comparability between the case studies through the common structure and format used for

each. Not only were these case studies comparable within the set; they were largely comparable with case studies from other studies conducted by the current team and their collaborators since they built on the established framework that had been used in previous studies.

One of the key challenges with the approach in this study was access to the relevant information. Since the research was conducted around 20–25 years ago, the recall of the researchers involved may not always have been perfect and relevant documentation may not have been available. A further limitation was the sample size because building these detailed case study narratives was resource intensive, and as such this limited the number of case studies that could be conducted within the study. This is turn had implications at the analysis stage, and made the case study selection process more important in terms of providing an appropriate spread of case studies across areas of interest.

Finally, a further limitation of the approach was that by focusing on what had happened, as a case study necessarily does, no direct data were collected on potential counterfactuals. It is not possible to assess formally what might have happened under different circumstances. What would have resulted from a piece of research if there had been a different approach to funding research? Or if the research had been conducted by a different research group? Or if peer-review processes had been changed in some way? It is not possible to say with any certainty whether any success (or lack thereof) would have occurred under different circumstances. We can only look at correlation – from which it is difficult to establish causation – and at the researchers' and other interviewees' insights into the process, which at times involved informal consideration of possible counterfactuals.

CHAPTER 6 Perspectives

6.1 Structure of the perspectives and approach

The aim of the perspectives was to provide a detailed view of the way in which some interventions of proven and potential value were developed, the research that underpinned them, and the process by which this research was translated into practice (or close to practice). The perspectives were included in the study because we were concerned that it might be challenging to identify research that had directly translated into health and broader economic benefits. If we did not have case studies where translation had occurred, it would prevent us from analysing the later steps of the translation process. The perspectives were intended to allow us to examine directly some of the more downstream elements of the research translation pathway, as well as illustrating some of the benefits that have resulted from research conducted in mental health over the last 20–25 years.

We use the term 'perspectives' to describe these narratives, since it became clear to us that progress in the field is often controversial, and it was beyond the resources of the project to identify a single clear narrative which was universally approved and agreed on to capture the understanding of all stakeholders. Therefore, each perspective offers a view on the developments in a particular area that reflects the balance of the many views of researchers and other parties consulted in the production of the narratives. As far as possible, we have constructed detailed and balanced accounts. However, we do not claim they offer a full and definitive picture of the development of a particular intervention or potential intervention as understood from all viewpoints. Therefore, they remain 'perspectives'.

As this was a new element in the Retrosight methodology, the structure for the perspectives was developed iteratively as the research was conducted. The final structure used for the perspectives is shown in Figure 6-1, along with the various aspects we considered in constructing them.

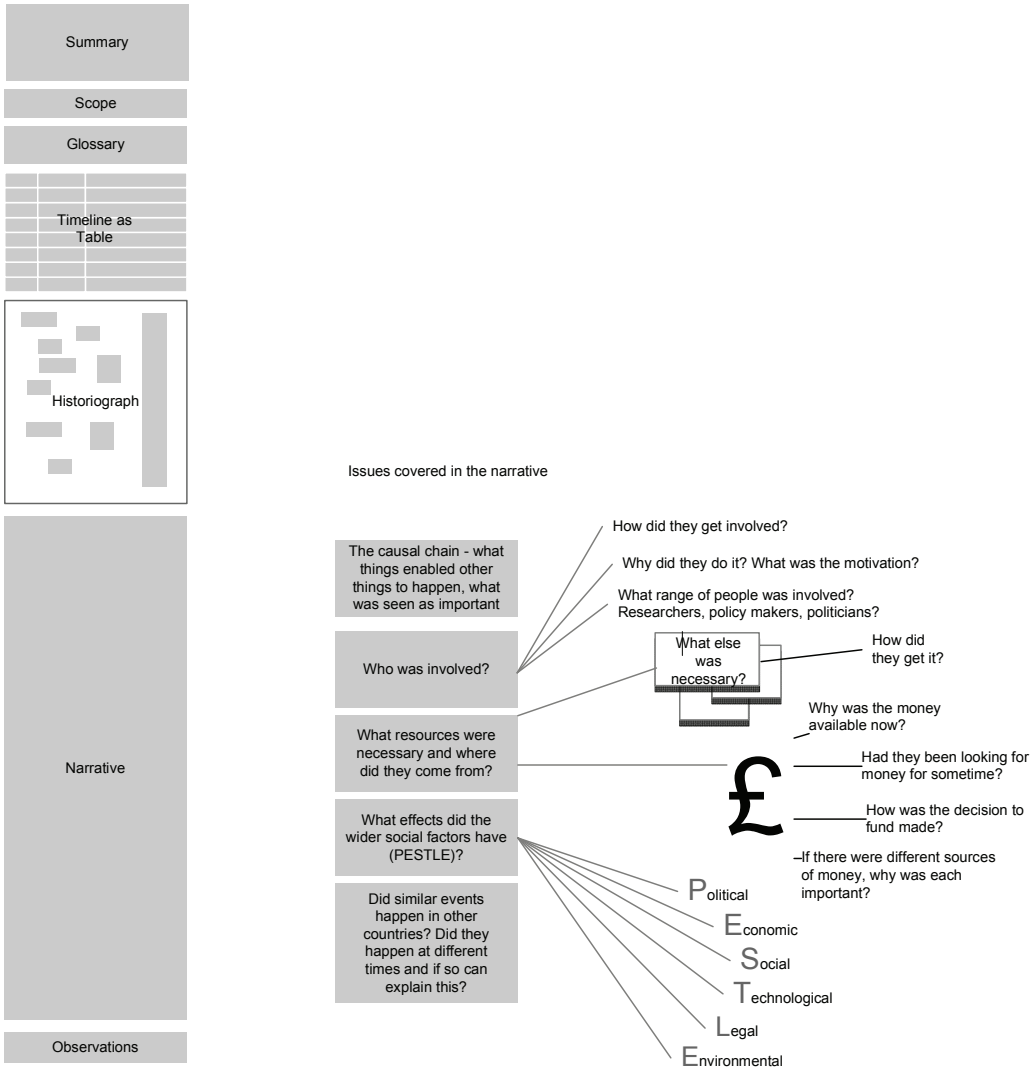


Figure 6-1 Structure of perspectives

The content and purpose of each of the main sections of the perspectives is as follows:

Summary: This provides an overview of the content of the perspective and acts as an abstract for the document.

Scope: In this section the overall scope of the perspective is defined. In many cases, interventions may be defined to include different elements and in different ways; for example, early intervention in schizophrenia may be defined to mean either intervention before the first episode, in the prodromal period, or immediately after the first episode – or both of these. This section clarifies the scope used for the perspective.

Timeline or table: In this section the key events and steps in the development of the intervention and the lines of research that contributed to it are laid out sequentially.

Historiograph: This is a graphical representation of similar information to that presented in the timeline or table. Events are set out against a timeline, but connections and interactions between events are indicated. Additionally, colouring is

used to represent the country in which particular developments occurred to show the transmission of research and implementation between countries.

Narrative: This provides a detailed account of the events surrounding the research development, translation and implementation. It covers the following:

- The causal chain: what things enabled other things to happen? What was seen as important?
- Who was involved, and what was their role?
- What resources were necessary and where did they come from?
- What effect did wider social, economic, political and systems factors have?
- Did similar events happen in different countries? Did they happen at different times and if so, how may that be explained?

Observations: This section highlights particular observations that the researchers carrying out the perspectives noted about the narrative.

6.2 Data collection

The perspectives were developed in a broadly similar way to the case studies, the main data sources being key informant interviews and desk-based research. However, the ordering was slightly different, reflecting the need first to understand and start to delineate the intervention or advance in order to identify relevant interview candidates. This delineation was done through a review of recent research articles and practice-related documents, such as guidelines, policy documents and training curricula. We aimed to explore definitions of the intervention, the extent of its uptake in the three project countries and the scope of its use, and to produce and outline the historical path of its development.

This initial review began an iterative process of interviews and further literature review as we built up the perspectives. The focus was on identifying key decision points in the development of the advance, and addressing questions about both its research antecedents and contextual facilitators and barriers in translation and adoption into practice. Therefore, desk research and interviews were interwoven throughout the development of the perspectives.

The approach used for interviews was open and a standardised interview protocol was not used, reflecting the differences between the types of advance explored and the issues that would need to be explored in interviews. However, in common with the case studies, the interviews were conducted confidentially and quotes were used only with the approval of the interviewees. Informed consent of all interviewees was sought in advance of the interview, and all interviewees were given the opportunity to review and comment upon the draft perspective at the validation stage. In total, 34 interviews were conducted across the six perspectives. All interviews were recorded, with the permission of the interviewee, and were transcribed. Interviews were conducted both in person and by telephone, depending on the level of input required from the interview, and the logistical feasibility of conducting interviews in person.

6.3 Validation

For the perspectives, the validation process consists of review and approval by the people interviewed previously. A further 'external' review was not sought because it became clear to us that it was not possible to generate a single agreed narrative that would be accepted by everyone. We tried to conduct sufficient interviews to allow a coherent narrative to emerge, with facts and views supported by multiple sources. At this point, we considered that we had a relatively balanced account of developments.

The perspectives were sent to all people interviewed for review and, where provided, their comments and suggestions were taken into account in revisions for the final version of the perspectives.

6.4 Strengths and weaknesses of the approach

One of the strengths of the perspectives is that they did succeed, as hoped, in documenting clear examples of the translation of research into advances in care for schizophrenia. Because of their backward-tracing approach, they also gave us access to information on the later stages of the translation pathway, which are more difficult to access through the case studies, and show how different pieces of research have come together to contribute to a particular development. The narratives produced were detailed, providing an in-depth understanding about the timeline through which developments took place across all three study countries and more widely.

The challenge again, as with the case studies, was that this was a resource-intensive approach, meaning that within the scope of the project it was possible to conduct only a limited number of perspectives, so this led to a limited dataset. Furthermore, since these perspectives were not based on an previously defined framework – such as that used in the case studies – there was a risk that they were less comparable with one another than the case studies, and that there may have been differences in content between them, though attempts were made to standardise the format to address this potential issue, as described above. Another potential weakness was that the narratives could not be considered authoritative. Developments in the field of mental health research have been controversial in some cases, so these descriptions remain perspectives, based on the views of the wide range of informants contacted, rather than a narrative that represents a consensus amongst key stakeholders.

CHAPTER 7 Analysis

The aims of the analysis phase were (i) to identify the long-term payback from mental health research, (ii) to identify factors that were associated with successful translation of research, and (iii) to draw out insights to inform future funding policy. The factors identified might be characteristics of the research itself (e.g. use of a new technique), the team carrying out the research (e.g. the motivation of the PI), or the environment in which the research took place (e.g. a clinical setting).

Achieving these aims involved a number of steps which are outlined in Figure 7-1. The first stage, to analyse the case studies, consisted of two main tasks: rating the case studies in terms of their impact, both within the research system and in society more widely; and identifying common factors or characteristics across the case studies. This allowed us to turn the extended case study narratives into data that could be analysed and compared systematically. The results of these two tasks were then combined, using a statistically guided narrative analysis approach, to examine the association between the various factors and impact.

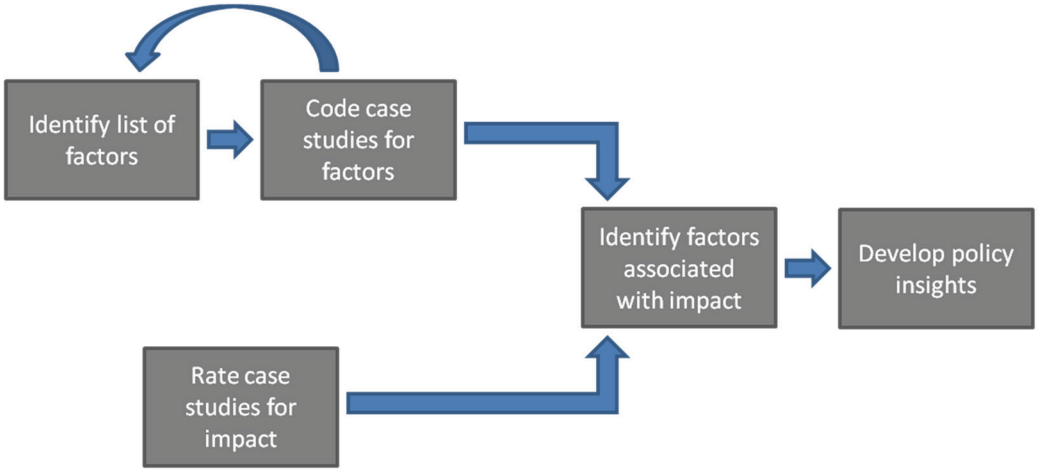


Figure 7-1 Outline of analysis phase

The analysis was complemented by a narrative analysis of the perspectives. This aimed to test some of the findings of the case study analysis, and also to draw out further insights and observations – particularly in relation to the more downstream elements of translation, which are better covered by the perspectives than the case studies.

We describe these analysis stages in turn in the sections below.

7.1 Coding the case studies for factors likely to influence impact

In this stage we identified characteristics and factors that might be associated with the level of impact achieved within a research cloud. To do this, we coded the case studies against a codebook using NVivo qualitative analysis software (QSR International Pty Ltd, 2010). The output of this process was twofold: a list of those case studies in which each factor was identified, and the text relating to the factor in each case study so that the information could be revisited in depth at the later analysis stage.

7.1.1 Identifying a provisional list of factors

In order to code the case studies, we identified a list of factors to serve as an initial codebook for researchers to work from. This consisted of a list of factors (some with subfactors) and definitions so they could be applied by multiple researchers. The initial list of factors came from a number of sources such as our previous projects in cardiovascular and arthritis research; others were suggested by the research team as the project progressed. The factors were organised into themes. Our provisional list of factors was piloted on a set of six case studies (two basic, two clinical, two interventional) and then further refined. The refinement process continued during the coding process, with regular meetings of the full team to discuss possible new factors emerging which could be added to the codebook.

The final codebook consisted of 49 factors grouped into six categories: researcher characteristics, funding characteristics, research characteristics, institutional characteristics, system characteristics, and other. The full list of codes used and their definitions is provided in Appendix C.

7.1.2 Coding the case studies

Coding was performed using the qualitative analysis software package NVivo (QSR International, 2010), which allowed us to keep a systematic record of the process and ensured that coding information was stored in a flexible format for analysis. We double coded each case study – that is, both the case study author and another member of the research team analysed each case study independently. One advantage of involving authors was that it highlighted any contextual information that the author might have picked up while developing the case study, but that is not clear in the narrative. After both coders had completed their coding, they met to discuss and resolve any coding disagreements. Broadly, we found that in this process there was little disagreement on the coding, but that around 25% of the time one researcher had identified a characteristic that the other agreed in discussion was present but had missed in their coding. On this basis, we assume that across the case studies we are more likely to have missed factors in case studies and 'undercoded' than to have incorrectly coded or 'overcoded'.

7.2 Scoring case studies for impact

In addition to identifying factors, the second key part of the analysis phase was to establish the level of impact of each research cloud. To do this, we used a scoring process in which the 18 case studies were scored against five impact categories (corresponding to those in the Payback Framework) by a scoring panel. The panel

consisted of experts drawn from the three project countries. These experts scored the case studies in isolation and then participated in a workshop where they discussed case studies, and finally they were given the opportunity to rescore case studies on the basis of those discussions. These stages, and the scores resulting from the process, are described below.

7.2.1 The scoring panel

The panel was selected to include a mix of researchers, practitioners, service-users and policymakers / research funders. We also wanted to include in the panel some people who were familiar with the study and the Retrosight methodology. Finally, we also wanted to include a balance of people from across the study countries. The final composition of the scoring panel is shown in Figure 7-2. This panel is more diverse than that used in previous studies. For example, in Cardiovascular Retrosight, the panel consisted primarily of members of the study team and policymakers/research funders.

	Canada	USA	UK	Other
Basic researcher / practitioner			*Roland Ryall*	
Clinical researcher / practitioner	*Rohan Ganguli*			
Interventional researcher / practitioner		*Ben Druss*		
Service user / family	*Ian Boeckh*		*David Jordan*	
Research policy maker / Funder / Evaluator	*Nathalie Gendron*	*Alan Kraut*		*Brendan Curran (Ireland)*
Continuity / Research team			*Jonathan Grant*	

Figure 7-2 Composition of scoring panel

7.2.2 The scoring process

Each member of the scoring panel rated each case study in each of five categories of impact (taken from the Payback Framework – Buxton & Hanney, 1996): knowledge production, research targeting and capacity building, informing policy and product

development, health and health sector benefits, and broader economic benefits. In addition, they were asked to provide a score for 'overall impact' for each case study.

Scorers were instructed to rank all the case studies on a scale between 0 and 9 for each category of impact. The scoring was within set – that is, the scores were effectively a ranking of the case studies within a particular category. Therefore, it is important to stress that the scores should not be directly compared – a 9 in one impact category is not equivalent to a 9 in another category. The specific instructions given to the panel regarding the scoring process were as follows:

1. Familiarise yourself with the case studies by reading over the summaries. You may also wish to review 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).

The panel were asked to score within set in this way in order to minimise the level of subjectivity in the scores – it is difficult to interpret scores which are relative to some 'ideal' impact outcome that is likely to differ between scorers.

The panel were sent a range of materials on the basis of which they conducted their initial round of scoring. These materials aimed to summarise the case studies to aid the scorers since reading and taking in the full content of 18 case studies for comparison is difficult. The materials provided were as follows:

– *Payback table:* Summarised key outputs of the case study under each of the categories listed above.

– *Case study summary table:* Summarised the case study narrative, organised by stage of the research process. Each section could be cross-referenced back to the full case study text, and gave a fuller explanation of the context than the payback table, with additional background material.

– *Bibliometrics data:* Key data for each case study presented as a series of comparative charts. More detailed bibliometric data were also provided for each individual case study.

In addition, the full case study text was provided for all case studies. We did not expect the scorers to read all the case studies in full (although through the discussions at the workshop it emerged that around half the panel had done this); rather, we expected

that the full text would be used for reference and clarification where necessary as a supplement to the summary data provided for scoring.

Scorers were given two weeks to review the materials provided and score all 18 case studies across the six categories (five payback categories plus overall impact). Scores were received from eight of the nine panel participants.

7.2.3 The workshop

Once initial ratings had been given, a workshop brought together the panel to discuss the case studies. The aim of this session was to reduce any differences in understanding of the research described in the case studies, but preserve any differences that remained in raters' impact scores that remained after discussion. After discussion of the case studies as a group, the panel were given the opportunity to rescore the case studies on the basis of this fuller shared understanding of their content. These ratings were then used for further analysis in the project.

The scores produced were subsequently analysed to look at the level of consensus between raters. This showed that there was a considerable level of consensus, sufficient to make comparisons between the high performing and low performing case studies in the different impact categories. The correlations between the different impact categories were also investigated, and it was found that there was correlation between most categories to differing extents, with adjacent impact categories typically most highly correlated. For example, research targeting and capacity building is correlated most strongly with knowledge production and informing policy and product development, and less well correlated with health and health sector benefit and broader economic benefit.

The overall impact category was found to be correlated with all the other categories, but through discussion at the workshop it became apparent that raters had reached these overall impact scores in different ways. Some scorers had just averaged scores across categories. Others had weighted more downstream impacts more heavily (e.g. broader economic benefits or health sector benefits), and some had weighted knowledge production more heavily. This lack of consensus about the relative weighting of different elements within the overall impact score led us to decide that this overall score would not be used in the later analysis.

Since all categories of impact are correlated to some extent and there is no obvious cut-off point or clear way to separate impact categories, impact is analysed in each category individually. To do this, a group of high- and low-impact case studies has been identified for each category, by ranking each case study in each category individually, and then dividing each category into thirds – or as close to that as possible – keeping tied rankings together. This generated a high-impact, low-impact and mid-impact group of case studies for each impact category. There are between five and seven case studies in each group. The analysis focuses on the high- and low-impact groups, comparing factors that are present or absent in each of these groups by category.

More detail about the workshop and the analysis of the scores from the workshop are presented in Appendix D.

7.3 Combining the data: a statistically guided narrative analysis approach

Having determined the level of impact and the factors occurring in each case study, we combined these two datasets in order to identify factors that were associated with either high or low impact in different categories, and to consider how likely it was that these associations would have occurred by chance. For example, if we found a particular factor was present in four of six high-impact case studies in a knowledge production, but only two of the six low-impact case studies – we could estimate, using the Fisher Exact Test, how likely this distribution is (Cochran, 1954). We employed Fisher's exact test to consider the role of chance in the associations observed in our data. This test is preferable to the chi-squared tests more commonly used in such settings when sample sizes are small, as they were here. It should be noted that both tests assumed that the data were a simple random sample from the population of interest. Because that assumption may not hold with these data, the results of these tests should be interpreted with some caution. We presented these tests results heuristically, as providing information on the relative, rather than absolute, strength of evidence of associations beyond what one would expect from chance alone. This gave us an initial way to filter the large amount of data for factors which were most likely to be of interest. We used an initial cut-off of less than 0.25 to select the factors for qualitative analysis. In addition, we also looked at factors where we didn't see any association with impact, but where from previous studies we would have expected to see a relationship.

The qualitative analysis of those factors identified for further investigation primarily consisted of revisiting the case study text coded to the relevant factor to understand in more detail why that case study had been coded to a particular factor, to understand the way the factor was interpreted in more depth, and to understand the relationship and possible implications in more detail. The overall process is laid out in Figure 7-3.

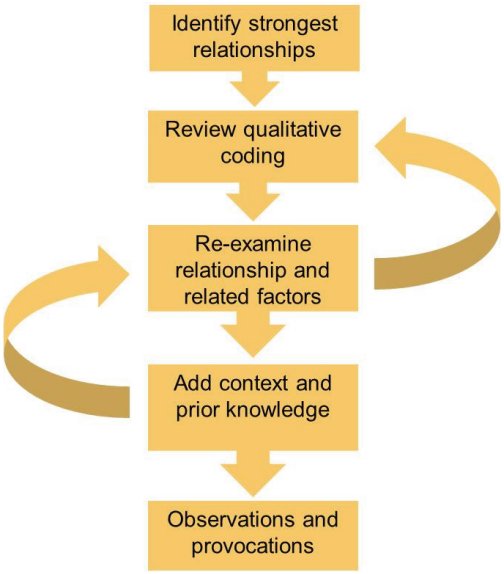


Figure 7-3 Statistically guided narrative analysis approach: overview

Broadly, our overall set of criteria for identifying factors of interest was as follows:

- *Strength of association:* How confident were we of the relationship between the factor and level of impact? We were more inclined to trust relationships where the association was less likely to be due to chance.
- *Causal plausibility:* Does the relationship have a credible causal chain, given our experience of research translation? If it occurs, can we understand how it might be true? We gave more weight to associations where there was a credible causal chain.
- *Previous research:* Does the relationship align with previous research on the science of science? We gave more weight to relationships that fitted with previous research – where that research was applicable to the setting we were examining. We also used previous research to identify facets of the relationships that we should probe more deeply, and other potential correlates or confounders that we should explore.
- *Attendant associations:* Did we see associations in neighbouring payback categories that we would expect, given the association in the category being examined?
- *Policy relevance:* Could we draw interesting – and more importantly, useful – policy insights from the association? We tended not to pursue further those relationships that would not lead to a policy-relevant finding. However, we did not select findings that were policy relevant where there was not already a strong association statistically and other evidence of plausibility.

This process allowed us to generate a shortlist of associations that appeared most robust and relevant to the aims of the project, to use to develop initial policy insights.

7.3.1 Developing policy insights

At an internal workshop we discussed the observations identified and started to group those which may be related or combined in terms of policy observations. After this clustering exercise, we again revisited the text of the case studies to look at these relationships in more detail and further develop the observations and policy insights. This was an iterative analysis process with the aim of drawing out implications for funding policy from the associations we have observed.

These initial findings were then tested in an 'emerging findings' workshop held in New York on 13 November 2012. The workshop was attended by a group of experts who were invited to critique our analysis, discuss any issues raised by it, contextualise our observations based on their experiences, and suggest further iterations of analysis or areas to examine in more detail. A list of the organisations represented at the meeting is provided in Box 2 below. Following the workshop, a further round of analysis was conducted to develop a final set of policy-relevant insights.

The Graham Boeckh Foundation (Canada)

University College London (UK)

Ontario Brain Institute (Canada)

Centre for Addiction and Mental Health (Canada)

Brain and Behaviour Research Foundation (USA)

Canadian Institutes of Health Research (Canada)

Columbia University (USA)

National Institute of Mental Health (USA)

Alberta Innovates – Health Solutions (Canada)

Douglas Institute (Canada)

Emory University (USA)

EJLB Foundation (Canada)

Wellcome Trust (UK)

University of Maryland (USA)

King's College London (UK)

Box 2 Organisations represented at the Emerging Findings Workshop

7.4 Examination of classification and possible confounding characteristics

In addition to the factors we identified, there were other characteristics of the case studies that might affect the level of impact. This section considers some of those possible confounders and the steps we took to test whether they were affecting the level of impact in the case studies.

7.4.1 The basic/clinical classification

One of the key classifications we used was to distinguish basic and clinical case studies (or more generically basic and applied research). Unfortunately, there are a number of different definitions for the two categories and the clinical/basic distinction is not necessarily mutually exclusive, so we tested a variety of ways to define different types of research.

We used both National Institutes of Health Research (NIH) and Frascati classifications. These classifications differ in how they classify research: the NIH clinical definition focuses on the research subject, while the NIH basic and Frascati definitions concentrate on the motivation for the research (full definitions are given in Table 7-1). The definitions also differ in the scope of research they encompass: NIH definition is intended to apply only to biomedical and health research, whereas the Frascati definition is intended for compiling international R&D expenditure comparisons across all fields.

We could not use the original definitions to select the initial research papers, because that system was tied to journal publication; we now wanted to classify bodies of research by their content.

Table 7-1 Definitions of different types of research

	NIH	Frascati
Basic	Systematic study directed towards fuller knowledge or understanding of the fundamental aspects of phenomena and of observable facts **without specific applications in mind.**	Experimental or theoretical work undertaken primarily to acquire new knowledge of the underlying foundation of phenomena and observable facts, **without any particular application or use in view.**
Clinical or applied research	*(1)* Patient-oriented research. Research conducted with human subjects (or on material of human origin such as tissues, specimens and cognitive phenomena) for which an investigator (or colleague) directly interacts with human subjects. **Excluded from this definition are *in vitro* studies that utilize human tissues that cannot be linked to a living individual**. Patient-oriented research includes: (a) mechanisms of human disease, (b) therapeutic interventions, (c) clinical trials, or (d) development of new technologies. *(2)* Epidemiologic and behavioral studies. *(3)* Outcomes research and health services research.	Original investigation undertaken in order to acquire new knowledge. It is, however, **directed primarily towards a specific practical aim or objective**.
Interventional or Experimental		Systematic work, drawing on existing knowledge gained from research and/or practical experience, which is directed to **producing new materials, products or devices, to installing new processes, systems and services, or to improving substantially those already produced or installed.**

Key phrases of longer definitions are in bold type.

The case studies were classified in a largely similar manner, regardless of which of the definitions applied – but there were some differences (Table 7-2). Importantly, both definitions gave us a spread of case studies across the types of research.

Table 7-2 Classification of case studies using different definitions of basic and applied research

NIH Clinical	Frascati
Basic	Basic
Basic	Basic
Basic	Basic
Neither	Applied
Both	Basic
Basic	Basic
Clinical	Experimental
Basic	Basic
Both	Basic
Clinical	Applied
Clinical	Applied
Clinical	Applied
Clinical	Experimental
Both	Basic
Clinical	Experimental
Both	Basic
Clinical	Experimental
Clinical	Experimental

Because the NIH definitions of clinical and basic research were not mutually exclusive, we ended up with three different classifications:

1. NIH inclusive, where we count any case study classified as both clinical and basic in both categories.
2. NIH exclusive, where we counted only case studies classified as clinical or basic.
3. The Frascati definition that applies across all areas of research, not just biomedical research.

In addition to calculating mean scores for each type of research basic/clinical comparison, we also examined the distribution of those scores. This is shown in Figure 7-4.

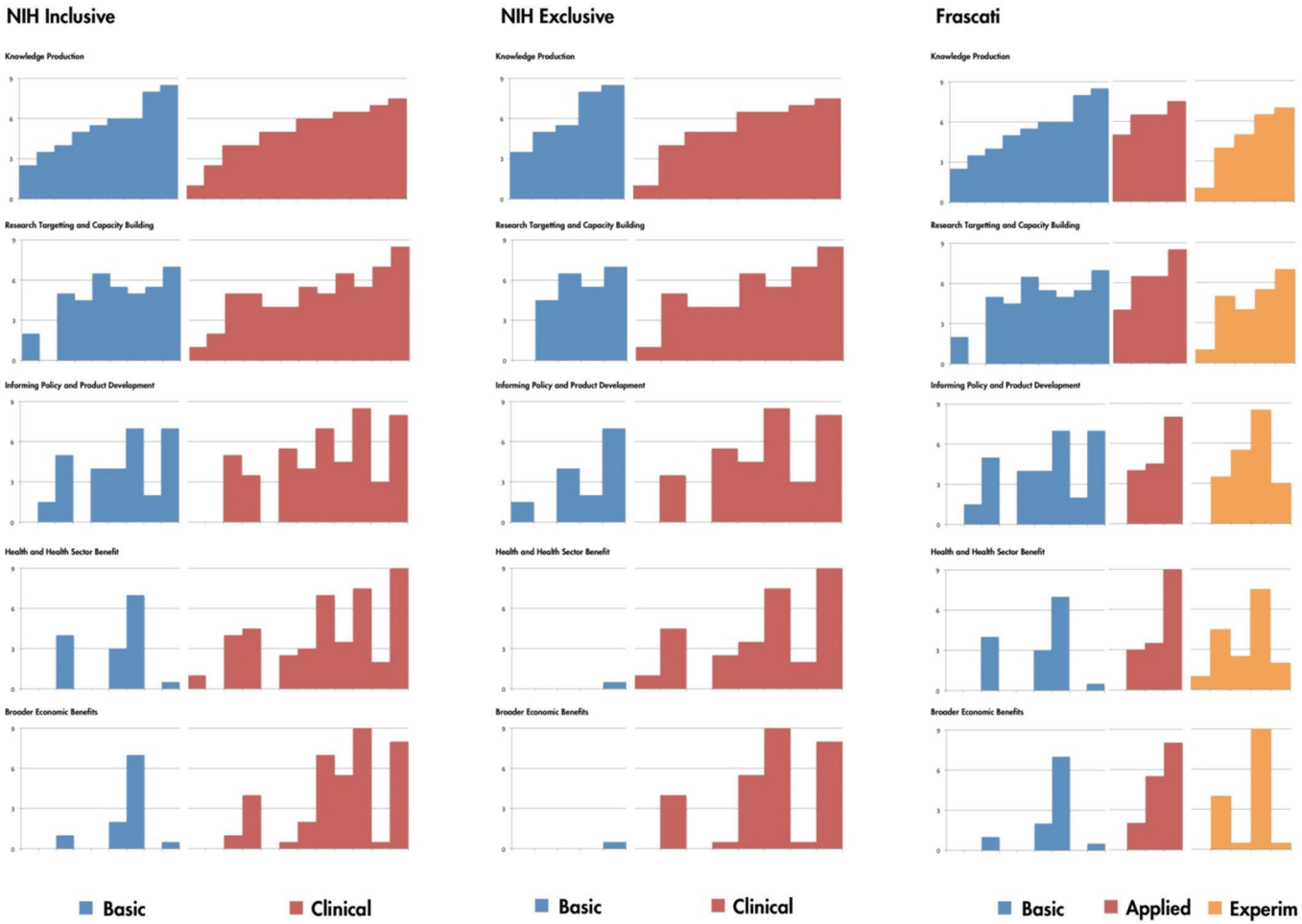


Figure 7.4: The distribution of impact scores by case study across the payback categories for the three classifications examined. For each classification the case studies are shown sorted in order of knowledge production impact score. For NIH Inclusive definition some case studies are counted in both basic and clinical hence more than 18 bars are shown, for NIH Exclusive case studies classified as both or neither as not included so only 13 bars are shown.

7.4.3 The effect of case study authorship

Six researchers wrote case studies, and this raised concerns about the consistency of examination and write-up. To strengthen the consistency, we used a standardised framework and protocols and had regular team meetings. We also assigned teams of two researchers to each case study and ensured that there were as many different pairings as possible across the full set of case studies. Finally all the case studies were reviewed and commented on by two researchers who were not authors of any of the case studies before undergoing validation by the interviewees, external peer review and review by one of the project's quality assurance reviewers. We cannot see any statistically significant relationship between the case study author and the average case study scores in the scoring exercise.

7.4.4 The effect of resources used by the research cloud on clinical/basic comparison

Because we examined research clouds, which often were supported by multiple funding sources over extended periods of time, we had no direct way to estimate the resources required for each piece of research. We collected information on the grants that researchers received, but it was impossible to quantify accurately the level of infrastructure and indirect costs that supported the work. To address this, we asked each member of the project team to assess pairs of case studies that they were involved in writing and suggest which was larger in terms of resources. We then combined that into an overall network which is shown in Figure 7-5. The consistency of comparison is high, although there are some disagreements, and it is clear from the network that clinical case studies are not consistently larger than the basic case studies.

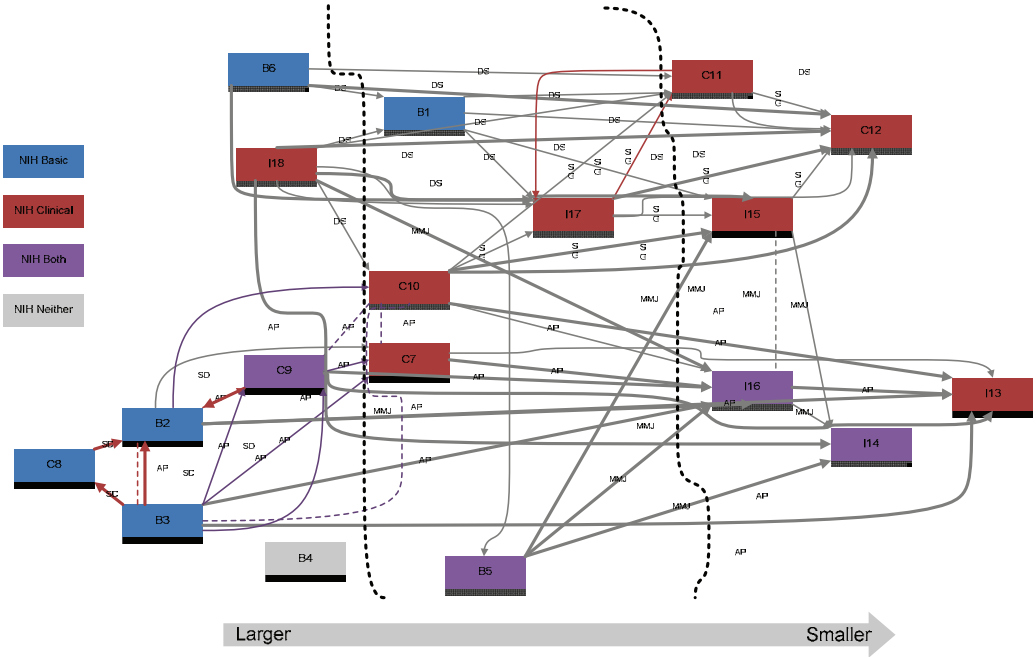


Figure 7-5 Overall network for pairwise resource comparisons of the case studies. Arrows indicate 'is larger than', thick arrows 'is much larger than', grey dotted lines 'is about the same size as'. Red lines indicate direct contradictions. The black dotted lines split the cases into large, medium and small grouping

7.4.5 The effect of research domain on clinical/basic comparison

We were concerned that different mixes of specific types of study within the basic and clinical general research categories might drive difference in impacts observed between these categories. To investigate this we broke down the case studies by research domain – 'Biological', 'Psychosocial' and 'Diagnostic'. These domains are defined in Table 7-3.

Table 7-3 Research domain definitions

Domain	Definition
Biological	relating to cell signalling, receptors, possible drug targets or drugs
Psychosocial	relating to psychosocial interventions or patient relationships and situation
Diagnostic	observational techniques and studies that could be applied to patients in no or any type of treatment (e.g. PET, identification of early signs, understanding co-morbidities)

We then looked within the 'Biological' domain to examine the differences in impact profiles as before, shown in **Error! Reference source not found.**. Despite narrowing the comparison to the 'Biological' domain only, the clinical case studies maintained the lead in health, social and economic impact that had been previously observed. Interestingly, for the two NIH classifications the basic case studies had a larger impact than the non-basic case studies in the knowledge production category.

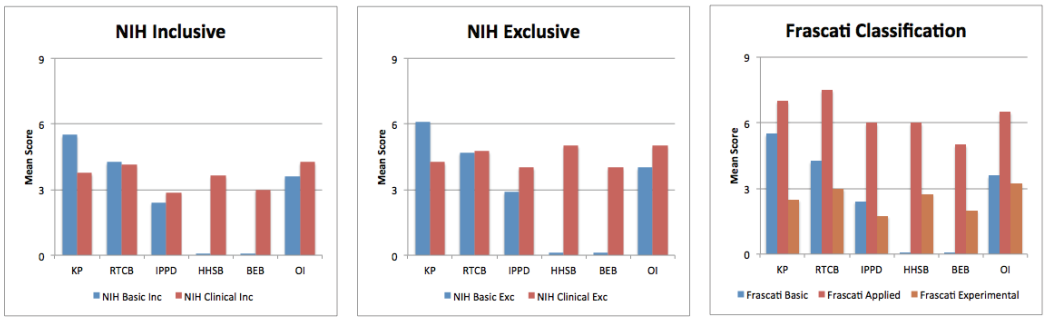


Figure 7-6 The mean impact scores of 'Biological' case studies for different types of research across the five impact categories

7.5 Analysis of perspectives

We carried out a thematic analysis of the six perspectives to assess the factors that contribute to, or hinder, developments in mental health research and the implementation of interventions. These factors were identified by two separate means. First, three members of the research team each coded two perspectives for a range of

factors, both those previously explored in the case studies and (through an iterative process) others that emerged during the course of the analysis. Secondly, a fourth member of the research team reviewed all six perspectives and drew out a series of common themes. The team as a whole then met to discuss their findings, and both analyses were combined in developing observations.

7.6 Strengths and weaknesses of the approach

A key strength of this analysis approach is the way it combines both qualitative and quantitative approaches. This allows for a detailed qualitative analysis of the content of the case studies but, by using the statistical analysis as an initial filter, increases the transparency of the analysis and reduces the opportunity for researcher-induced biases. The iterative, statistically guided narrative analysis approach allows us to develop, test and refine findings, to ensure that they are robust and credible.

A second strength of the approach is that we have multiple success measures – case studies are judged on five axes of success from knowledge production to economic benefit – allowing us a nuanced view of the different factors likely to contribute to different types of impact. The importance of breaking down the concept of 'impact' in this way was further demonstrated by the very different opinions of our expert panel on what 'overall' impact should represent. In some cases, scorers considered a high overall impact to be dependent on generating wider societal benefits, while in others a higher value was placed on advancing knowledge.

In terms of the initial analysis of the scoring data, the approach is transparent and the level of agreement between scorers indicates that it is possible to reach a consensus on the level of impact of the case studies across a range of categories. However, as the panel is relatively small the scores are based on the collective judgement of only a small number of experts. Given the diversity in the panel we have some confidence that the judgement is generalisable. but a different panel would have produced scores that were slightly different and might have reached a different conclusion on the relative merits of the case studies.

In terms of the coding process, we noted that there are likely to be some instances of missing coding, although we believe that is likely to be random and not to be biased for or against particular associations. This is inferred from the fact that when the coding was discussed between the two researchers involved, often one researcher had missed some instances that the other had captured, but both agreed that the coding was appropriate in discussion. On this basis, it is reasonable to assume that there will be some instances where something should have been coded that both researchers missed, but there is no reason to expect that there will be any particular biases. Because of the iterative approach of returning to the actual content of the case studies coded against specific factors, we will have corrected for data that were initially incorrectly coded, and instances of 'overcoding'.

The small number of case studies means that we could not expect to see very strong associations in our dataset and that, combined with the large set of factors (i.e. the possibility of many associations), is a weakness of the initial statistical filter; with so

many data, and a small number of samples, it is difficult to distinguish true relationships from noise. This is why we use these calculations only as an initial screening, followed by in-depth qualitative analysis of the case studies to understand the nature and credibility of the relationships suggested.

The use of 'high impact' and 'low impact' case study groups for the analysis, rather than an alternative approach (e.g. regression analysis), has advantages and disadvantages. The main disadvantage is the loss of data in what is already a small sample of case studies, since the data on the case studies in the 'mid-impact' group are not used. However, the strength of this approach lies in what we find in discussion with scorers (both at this workshop and from previous studies). Scorers typically find placing the best and worst case studies in a particular category relatively straightforward. However, placing some of the case studies with a moderate level of impact on the scale between these is more difficult. This partly reflects the approach we ask scorers to take in the scoring process. Therefore, by using only the high- and low-impact case studies in each category, the data we are using for the analysis are likely more reliable.

References

Buxton, M., & Hanney, S. (1996) How can payback from health services research be assessed? *Journal of Health Services Research and Policy,* 1, 35–43.

Grant, J., & Wooding, S. (2010). *In search of the Holy Grail: understanding research success.* Cambridge, UK: RAND Europe.

Hamilton, K. (2003) *Subfield and Level Classification of Journals* (No. 2012-R). Cherry Hill, NJ: CHI Research.

Hanney et al. (2006) Identifying the impact of diabetes research. *Diabetic Medicine*, 23, 176–184

Hanney, S., Gonzalez-Block, M., Buxton, M., & Kogan, M. (2003) The utilisation of health research in policy making: concepts, examples and methods of assessment. *Health Research Policy Systems*, 1, 1, 2.

Health Economics Research Group, Office of Health Economics, RAND Europe (2008) *Medical Research: What's it worth? Estimating the economic benefits from medical research in the UK.* London: UK Evaluation Forum.

IIT Research Institute and National Science Foundation (US) (1968) *Technology in Retrospect and Critical Events in Science*, vol. 2. Chicago: Illinois Institute of Technology Research Institute.

Insel, T.R. (2009) Translating scientific opportunity into public health impact: a strategic plan for research on mental illness. *Archives of General Psychiatry,* 66(2), 128–133.

Ismail, S., Nason, E., Marjanovic, S., & Grant, J. (2009). *Bibliometrics as a Tool for supporting Prospective R&D Decision-making in the Health Sciences: strengths, weaknesses and options for future development*. Cambridge, UK: RAND Europe.

Ismail, S. (2009) Delphi exercises. *Performance Audit Handbook*. Cambridge: RAND Europe, p. 27.

Kwan, P., Johnston, J., Fung, A.Y.K., Chong, D.S.Y., Collins, R.A., & Lo. S.V. (2007) A systematic evaluation of payback of publicly funded health and health services research in Hong Kong. *BMC Health Services Research,* 7, 121.

Levitt, R. Celia, C., Diepeveen, S., Ni Chonaill, S., Rabinovich, L., & Tiessen, J. (2010) *Assessing the Impact of Arts and Humanities Research at the University of Cambridge.* Cambridge, UK: RAND Europe, TR-816-CUAHRC, http://www.rand.org/pubs/technical_reports/TR816.

Narin, F., Pinski, G., & Gee, H.H. (1976). Structure of the biomedical literature. *Journal of the American Society for Information Science*, 27, 25–45.

Nason, E., Janta, B., Hastings, G., Hanney, S., O'Driscoll, M., & Wooding, S. (2008) *Health Research – making an Impact. The economic and social benefits of HRB-funded research.* Dublin: Health Research Board.

Oortwijn, W., Hanney, S., Ligtvoet, A., Hoorens, S., Wooding, S., Grant J. et al. (2008) Measuring the impact of health technology assessment in the Netherlands. *International Journal of Technology Assessment in Health Care*, 24, 259–69.

Panel on Return on Investment in Health Research (2009) *Making an Impact: a preferred framework and indicators to measure returns on investment in health research*. Canadian Academy of Health Sciences, Ottawa, ON, Canada.

Pollitt, A., Wooding, S., Hanney, S., Buxton, M., & Grant, J. (2011) *Project Retrosight: understanding the returns from cardiovascular and stroke research. Methodology report* (No. TR-925-RS). Cambridge, UK: RAND Europe.

Sherwin, C.W., & Isenson, R.S. (1967) Project Hindsight. A defense department study on the utility of research. *Science*, 23, 156(3782), 1571-1577.

Slote Morris, Z., Wooding, S., & Grant, J. (2011) The answer is 17 years, what is the question: understanding time lags in translational research. *Journal of the Royal Society of Medicine*, 104, 510–520.

Thomson Reuters. Web of Science. Retrieved 23 September 2011, from Thomson Reuters: http://thomsonreuters.com/products_services/science/science_products/a-z/web_of_science

Wooding, S., Hanney, S., Buxton, M., & Grant, J. (2004) *The Returns from Arthritis Research. Vol. 1: Approach, Analysis and Recommendations* (No. MG-251-ARC). Cambridge, UK: RAND Europe.

Wooding, S., Hanney, S., Buxton, M., & Grant, J. (2005) Payback arising from research funding: an evaluation of the Arthritis Research Campaign. *Rheumatology,* 44, 1145–1156.

Wooding, S., Keyser, D., Ni Chonaill, S., Schultz, D., Pincus. H., Pollitt, A., Horvitz-Lennon, M., Yaqub, O., Morris, Z., & Grant, J. (2010) *Mental Health Retrosight: final report on Phase I*. WR-789-GBF. UK: RAND Europe.

Wooding, S., Hanney, S., Pollitt, A., Buxton, M., & Grant, J. (2011) *Project Retrosight. Understanding the returns from cardiovascular and stroke research: policy report* (No. MG-1079-RS). Cambridge, UK: RAND Europe.

Appendix A: Details of bibliometric approach for selecting case studies through highly cited papers

This appendix gives a detailed description of the selection process for case studies which were identified through a bibliometrics exercise looking at highly cited papers published between 1985 and 1990. An overview of the approach is shown in Figure A-1 below, and the description that follows describes each stage shown in the diagram from left to right.

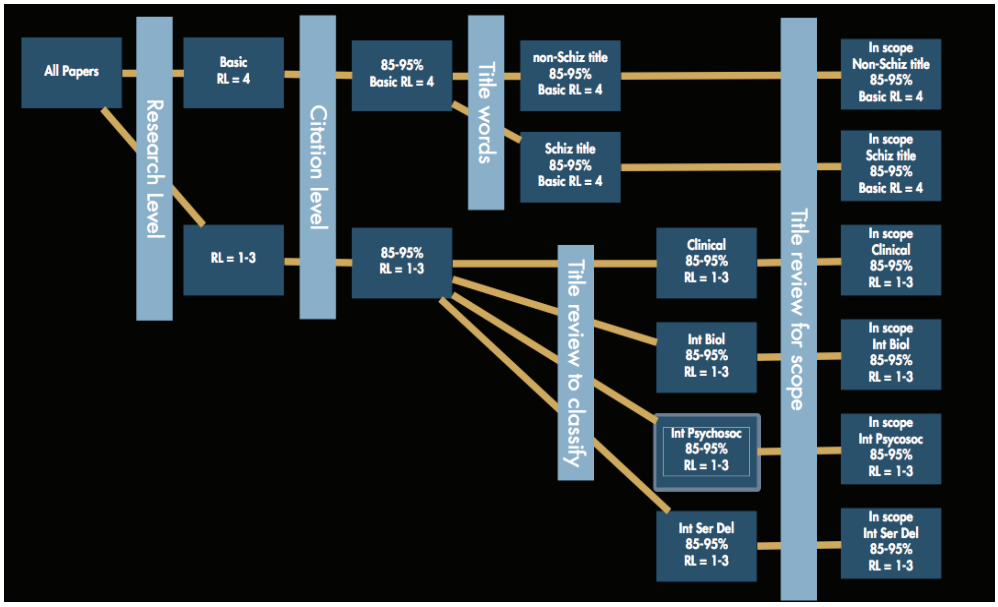


Figure A-1 Schema of process to identify hot research topics using bibliometrics

Identification of journal set

All neuroscience and mental health research papers with a US, UK or Canadian address published between 1985 and 1990 were identified from the Web of Science. The journal set was identified as:

- all papers published in 'Psychiatry', 'Neurology' and 'Neurosurgery' journals as defined by the National Science Foundation, and all papers in 'Psychiatry', 'Neuroscience', 'Clinical Neurology' and 'Neuroimaging' journals as defined by Thomson Reuters;
- all papers published in journals in which 75% or more of papers were retrieved using relevant MeSH terms[5] as summarised in Box 3;
- all papers published in other journals but retrieved using the relevant MeSH terms in Box 3.

This resulted in a dataset of 238,836 papers.

Mental Disorders [F03], except Substance-related Disorders [F03.900]

Mental Health Services [F04.408]

Mental Health [F02.418]

Neurosciences [H01.158.610]

Nervous System Diseases [C10]

Nervous System [A08]

Mental Disorders [F03]

Neurology [H02.403.600]

Neuropharmacology [H02.628.280]

Psychopharmacology [H02.628.546]

Box 3 MeSH terms used to identify highly cited papers

Grouping by research type

To facilitate the classification of papers by research type and to control for known differences in citation behaviour, we classified the papers using the National Science Foundation's 'Research Level' (Narin et al., 1976). This is a classification system that groups journals as being 'basic' (RL=4), 'clinical investigation' (RL=3), 'clinical mix'

[5] All of the articles indexed in MEDLINE have been read by an indexer. Indexers read each article in order to identify and assign the subject terms, from a standardised list, that describe the content of an article. MeSH is the acronym for this master list of standardised **Me**dical **S**ubject **H**eadings. Usually between 10 and 15 MeSH terms are assigned to an article to describe its content. OST maps the MeSH terms from MEDLINE on to its in-house bibliometric database derived from the Web of Science.

(RL=2) or 'clinical observation' (RL=1). At the outset we were aware that this system was potentially too crude given the unique breadth of mental health research, and this was confirmed by the very small number of schizophrenia-related papers occurring in our initial sample. However, we felt it would be a useful starting point that could be refined when reviewing the scope. We created two groups: 'basic research' (based on RL=4) and 'other' (based on RL = 1, 2 and 3). As explained below, the other category was then allocated into two other groups: 'clinical' and 'interventional' (with 'interventional' divided into three subgroups: 'biological', 'psychosocial' and 'health services'). The basic group had 73,310 papers and the other group 147,604. We excluded papers where the journal had not been classified (n=17,471).[6]

It should be noted that although we selected a range of pieces of research across these research types, this categorisation was applied only to the initial paper selected for each case study. In some instances this meant that the research cloud as a whole did not fit into the same category, and we addressed this by reclassifying the final case studies before analysis. However, the key aim of this approach was to ensure that we selected a suitable range of research and not to create a rigid system of classifying case studies.

Identification of highly cited papers

For the two groups of papers – basic and other – we identified the papers that fell between the 85th and 95th percentiles in terms of citations. We chose this range for the following reasons:

1. We wanted to identify 'hot' research – that is, we wanted papers that were highly cited in the five years following publication.
2. We needed to ensure that we could select at least two papers from Alberta, as AIHS was supporting two case studies.
3. We wanted to exclude outliers, and so not select the most highly cited.
4. We needed to ensure that we were comparing like with like in our paper selection and thus wanted a relatively tight percentile range.

We calculated the citation boundaries for the basic papers and other papers separately, as we knew from previous work that basic research has different citation patterns from non-basic research. We also explored the variance in the citation boundaries by the other three research levels (1, 2, and 3) to ensure that it was appropriate to put them in one group in the first instance. As shown in Table A-1, the 85th to 95th percentile range for basic research was between 29 and 61 citations, while for the other group it was 16 to 33 citations. Filtering papers based on these ranges resulted in 7,584 papers in the basic group and 7,183 papers in the other group.

[6] We explored the characteristics of these papers and found they were very lowly cited: the top cited paper in the group had 145 citations, but the median number of citations was 1 and the 90th percentile was 7 citations. This showed that we were not omitting many important highly cited papers. The reason that the journals were unclassified is that they had low visibility in the first place and thus publications in those journals are unlikely to be highly cited.

Table A-1 Citation boundaries for different research levels

		Citation boundary	
	RL	85%	95%
Basic	4	29	61
Other	1, 2, 3	16	32
	1	12	26
	2	16	33
	3	20	39
	NULL	5	11

Grouping by country

Using the address information of the first author, we then classified the papers to the USA, the UK and Canada. For Canada we also identified a subset of papers with an Alberta address to ensure that we identified two case studies from this province (given the support of AIHS). The distribution of papers by research group and country of corresponding author is provided in Table A-2 below.

Table A-2 Distribution of papers across countries

	Basic (RL =4)	Other (RL = 1, 2 and 3)
Canada (Alberta)	216 (31)	434 (22)
UK	392	1,006
USA	2,641	5,591
Total	3,249	7,031

Checking whether papers were 'in scope'

We were aware from the outset that the journal set was very broadly defined. We therefore wanted to filter the papers further to ensure that they were within the scope of the project. For the 'other' group we did this by simply filtering on 'schiz*' in the title of the paper – that is, we aimed to select papers that were explicitly focused on schizophrenia (and we subsequently removed papers focused on, e.g., schizotypal personality and schizophreniform disorder).[7]

We decided to split the basic group into papers that had a focus on schizophrenia and papers that did not (while potentially still being relevant to schizophrenia), and so again we used the presence or absence of 'schiz' in the title to split the set into two groups. We did

[7] We experimented with including the word 'psychosis' as a filter for the non-basic papers in addition to 'schiz*', but concluded that this did not add any additional specificity to our filtering.

this because one of the interesting findings from our work in cardiovascular research (Wooding et al., 2011) is that those basic research projects that had an explicit clinical motivation tended to have had a higher academic and wider impact. This was a potentially important observation that we wanted to explore further.

Table A-3 Distribution of papers by research level and country of corresponding author

	Basic (RL =4) no 'schiz*'	Basic (RL =4) with 'schiz*'	Other (RL = 1, 2 and 3) with 'schiz*'
Canada (Alberta)	216 (31)	0 (0)	11 (0)
UK	390	2	32
USA	2,637	4	170
Total	3,243	6	213

As is apparent in Table A-3, at this stage of the filtering process there was an uneven distribution of papers across countries and research categories. In some instances we had no candidates (e.g. Canada basic with 'schiz*' in the title); in others a shortlist of candidates; and in yet others a longlist that would need further filtering. The distribution of papers across the cells of the matrix made some sampling decisions for us. For example:

- the two Canadian basic research case studies had to come from 31 papers from Alberta that did not mention 'schiz*' in the title;
- we had enough papers to shortlist the UK (n=2) and US (n=4) basic research case studies with a mention of 'schiz*' in the title.

This categorisation led to two additional selection tasks:

1. to ensure that the basic research papers without 'schiz*' in the title for Alberta, the UK, and the USA were in scope; and
2. to allocate the other research papers for Canada, the UK and the USA into clinical and interventional subgroups.

Both tasks required the reviewing of titles. This was done by two members of the team (AP and SW) working independently; disagreements were resolved by discussion. As we applied these classifications, it became apparent that we needed to adjust our selection rules to ensure that we had enough candidate papers. These adjustments are described below and detail how we completed the final selection process for basic, clinical and interventional papers.

For all the papers selected we approached the corresponding author (or the first author if no corresponding author was identified) to ask them to participate in the study.

Final selection of six basic research papers

Basic papers without 'schiz*' in the title

Three of the 31 basic papers from Alberta without 'schiz*' in the title were deemed to be in scope by our researchers. However, two of these papers were written by the same research group and reported on the same research cloud. We selected the single paper and the more highly cited of the two by the same group:

- CHANGES IN STRIATAL DOPAMINE NEUROTRANSMISSION ASSESSED WITH MICRODIALYSIS FOLLOWING RECOVERY FROM A BILATERAL 6-OHDA LESION – VARIATION AS A FUNCTION OF LESION SIZE
- SYNAPTIC MODULATION BY DOPAMINE OF CALCIUM CURRENTS IN RAT PARS-INTERMEDIA

For these two papers we tracked down the relevant author through web searches, and asked the CEO of AIHS to approach them about participation in the study. Both researchers agreed to take part.

To reduce the number of papers to be reviewed for the basic with 'schiz*', for the UK and USA we focused on those papers at the 90th citation percentile – that is, the middle of our 85th to 95th percentile range. This reduced the shortlist to 17 papers from the UK and 106 papers from the USA.

For the USA, our reviewers initially identified 4 of the 106 US basic papers without 'schiz*' in the title as being potentially in scope. However, as we explain below, we could identify only one basic UK paper *with* 'schiz*' in the title, which meant that in order to have two UK basic papers we had to select one without 'schiz*' in the title. This prevented us from selecting any US papers without 'schiz*' (in order to ensure that we had three with and three without 'schiz*').

For the UK, 13 of the 17 papers were judged to be out of scope by our two reviewers. We contacted the first/corresponding authors for the four remaining papers, one of whom declined to participate. For the three remaining papers we looked at the incidence of the words 'schizophrenia' or 'psychosis' in the title of citing papers (i.e. those papers that used the work) and selected the only paper that met this criterion (with 3 out of 116 papers including either 'schizophrenia' or 'psychosis'). We made contact with the corresponding author, but he failed to respond to repeated communications and we were forced to reselect the case study. As there was no obvious reason to favour either of the two remaining papers, we tossed a coin. This resulted in our selecting the following paper:

- DIFFERENTIAL DISTRIBUTION OF $GABA_A$ RECEPTOR MESSENGER-RNAS IN BOVINE CEREBELLUM – LOCALIZATION OF ALPHA-2 MESSENGER-RNA IN BERGMANN GLIA LAYER

Basic papers with 'schiz*' in the title

As shown in Table A-3, there were no Canadian/Albertan basic papers with 'schiz*' in the title. Hence our three basic papers with 'schiz*' in the title had to come from either the USA or the UK.

We originally identified two UK basic papers with 'schiz*' in the title that fitted within our 85th–95th percentile range (which equated to 29–61 citations). Unfortunately, one of the papers was a review and therefore was excluded, and the first/corresponding author of the second paper had died. We then selected the next paper closest to our citation boundary (which had 27 citations). However, in this instance we were unable to locate the first/corresponding author. We therefore identified the next two papers outside the citation boundary (one above the boundary with 102 citations, the other below with 8 citations) and approached the first/corresponding authors. Both accepted, but it was then discovered that one was a co-author on another paper subsequently selected and therefore was excluded. The remaining paper, which was the less cited of the two, was selected. An interview was set up with the corresponding author of the selected paper, but before the interview could take place he withdrew from the project (due to time constraints).

At this point the project was too far advanced to reselect in this category without delaying the overall timescale of the project, so we elected to include one of the case studies from the project pilot. This case study was deemed a suitable replacement, as it was based around a UK basic paper (albeit without 'schiz*' in the title) and would allow us to continue the project with the sample size originally planned. The paper initially selected for this case study would not have fallen into our citation range for selection in the main project (it is more highly cited), but we felt it was the best option available to us.

The pilot case study taken forward into the main phase of the project was:

- IDENTIFICATION AND DISTRIBUTION OF 5-HT_3 RECEPTORS IN RAT BRAIN USING RADIOLIGAND BINDING

We identified four US basic papers with 'schiz*' in the title within our 85th–95th percentile citation range. One of these papers was excluded as it was a review. We identified and approached the corresponding/first author of the three remaining papers with the aim of recruiting two to the study. One of the three declined to participate, and so the two remaining, both of whom agreed to be interviewed, were selected as our US basic case studies.

- DYSFUNCTION IN A PREFRONTAL SUBSTRATE OF SUSTAINED ATTENTION IN SCHIZOPHRENIA
- ORGANIZATION OF DOPAMINE D1 AND D2 RECEPTORS IN HUMAN STRIATUM – RECEPTOR AUTORADIOGRAPHIC STUDIES IN HUNTINGTONS-DISEASE AND SCHIZOPHRENIA

Final selection of six clinical research papers

The 213 'other' papers (Table A-3) were then reviewed and allocated to one of four categories: clinical, interventional–biological, interventional–psychosocial and interventional–health services. Our reviewers agreed on the allocation of all the Canadian papers and all but one of the UK papers. For the 170 US papers they reached agreement on 145. We considered only the 187 papers where they had reached agreement.

The reviewers identified 10 (out of 11) papers in Canada as clinical. Five of these were disregarded as they were reviews, while the corresponding/first author had died for a

further two of the papers. This left us with three papers from which we had to select two. Two of these papers shared an author, so the third paper was selected automatically. Our Canadian sponsor selected between the two papers sharing an author. The selected author accepted following an initial invitation from our sponsor, but the corresponding author on the 'third' paper declined to take part. As there were no more suitable papers within our citation range, we selected the next paper above and the next below the 85th–95th percentile range. We were unable to trace the corresponding author for the more lowly cited paper, but the author of the more highly cited paper agreed to participate, resulting in the selection of the two Canadian clinical papers:

- AUDITORY P300 IN BORDERLINE PERSONALITY-DISORDER AND SCHIZOPHRENIA
- HUMAN-BRAIN D1 AND D2 DOPAMINE-RECEPTORS IN SCHIZOPHRENIA, ALZHEIMERS, PARKINSONS, AND HUNTINGTONS DISEASES

For the UK, of the 31 papers (out of 32) with reviewer agreement, 27 were classified as clinical. Of these 4 were reviews, leaving 23 papers from which 2 had to be selected. These papers were each assigned a random number and ordered from low to high according to this. Beginning with the first two papers on the list, we checked for duplication of institution or co-authors. The second paper selected was deemed out of scope (focusing on schizotypal personality traits, not schizophrenia) and the third shared an author with a previously selected paper. This resulted in our contacting the corresponding/first authors for the first and fourth papers on the list, both of whom agreed to participate. The papers selected were:

- RELATIVES EXPRESSED EMOTION AND THE COURSE OF SCHIZOPHRENIA IN CHANDIGARH – A 2-YEAR FOLLOW-UP OF A 1ST-CONTACT SAMPLE
- FURTHER INVESTIGATION OF THE PREDICTORS OF OUTCOME FOLLOWING FIRST SCHIZOPHRENIC EPISODES

For the 145 US papers with reviewer agreement, 116 were classified as being clinical. Of these 26 were reviews, leaving 90 papers from which 2 had to be selected. This was done in the same way as the selection of the UK clinical papers, through random ordering and exclusion of duplicate authors. In this instance the first and second papers on the list shared an author, and so we selected the first and third as our preferred papers. Both corresponding authors agreed to take part in the study, resulting in the selection of our two US clinical papers:

- LOW FRONTAL GLUCOSE-UTILIZATION IN CHRONIC-SCHIZOPHRENIA – A REPLICATION STUDY
- PREDICTION OF ADULT-ONSET SCHIZOPHRENIA FROM CHILDHOOD HOME MOVIES OF THE PATIENTS

Final selection of six interventional research papers

Of 187 (out of 213) 'other' papers where our reviewers agreed on classification, 34 were interventional; 26 of these 34 were biological, 5 psychosocial and 3 health services (Table A-4). We aimed to have two papers from each of the three interventional categories, spread across the three countries (i.e. two papers per country). The small numbers both helped and hindered selection.

Table A-4 Distribution of interventional research by type and country of corresponding author

	Biological	Psychosocial	Health services	Total
Canada	1			1
UK	2	1	1	4
USA	23	4	2	29
Total	26	5	3	34

As there were no psychosocial or health services papers in Canada, these needed to come from the UK and the USA and, given that there was only one paper in each of these categories in the UK, these two papers were self-selecting:

- SPECIFIC AND NONSPECIFIC EFFECTS OF EDUCATIONAL INTERVENTION WITH FAMILIES LIVING WITH A SCHIZOPHRENIC RELATIVE
- PSYCHIATRIC MORBIDITY OF A LONG STAY HOSPITAL POPULATION WITH CHRONIC-SCHIZOPHRENIA AND IMPLICATIONS FOR FUTURE COMMUNITY CARE

This meant that the other four papers (two biological, one psychosocial, one health services) had to come from Canada and the USA. For Canada the first/corresponding author on the sole biological paper had died, meaning that we had to relax the boundaries of our 16–32 citation range to identify two biological papers. The first of these (with 69 citations) we determined to be out of scope when we reviewed the paper. The second (with 15 citations) was in scope and selected. The next paper (with 13 citations) was also in scope and selected. Our Canadian sponsor approached the corresponding/first authors of the two papers, and both agreed to participate in the study. The papers selected were:

- VERAPAMIL IN THE TREATMENT OF CHRONIC-SCHIZOPHRENIA
- NEUROLEPTICS REVERSE ATTENTION ASYMMETRIES IN SCHIZOPHRENIC-PATIENTS

By choosing two Canadian biological papers, we were restricted in our choice of US papers to one from the psychosocial category and one from the health services category. Of the four candidate psychosocial papers, one was excluded as the corresponding/first author had died. We approached the corresponding/first author of the remaining three papers. One declined to take part and one we were unable to contact, meaning that the third, who agreed to participate, was automatically selected. Of the two health services papers, one was

a review and thus was excluded, while the corresponding/first author of the other paper agreed to take part. Thus we selected our final two interventional papers:

- GENDER AND SCHIZOPHRENIA OUTCOME – A CLINICAL-TRIAL OF AN INPATIENT FAMILY INTERVENTION
- ACUTE TREATMENT RESPONSE AND SHORT-TERM OUTCOME IN SCHIZOPHRENIA – 1ST RESULTS OF THE NIMH TREATMENT STRATEGIES IN SCHIZOPHRENIA STUDY

The final selection and distribution of the 18 papers across country and classification is illustrated in Figure A-2. In this matrix we include the title of the paper and the number of citations in the five years following publication.

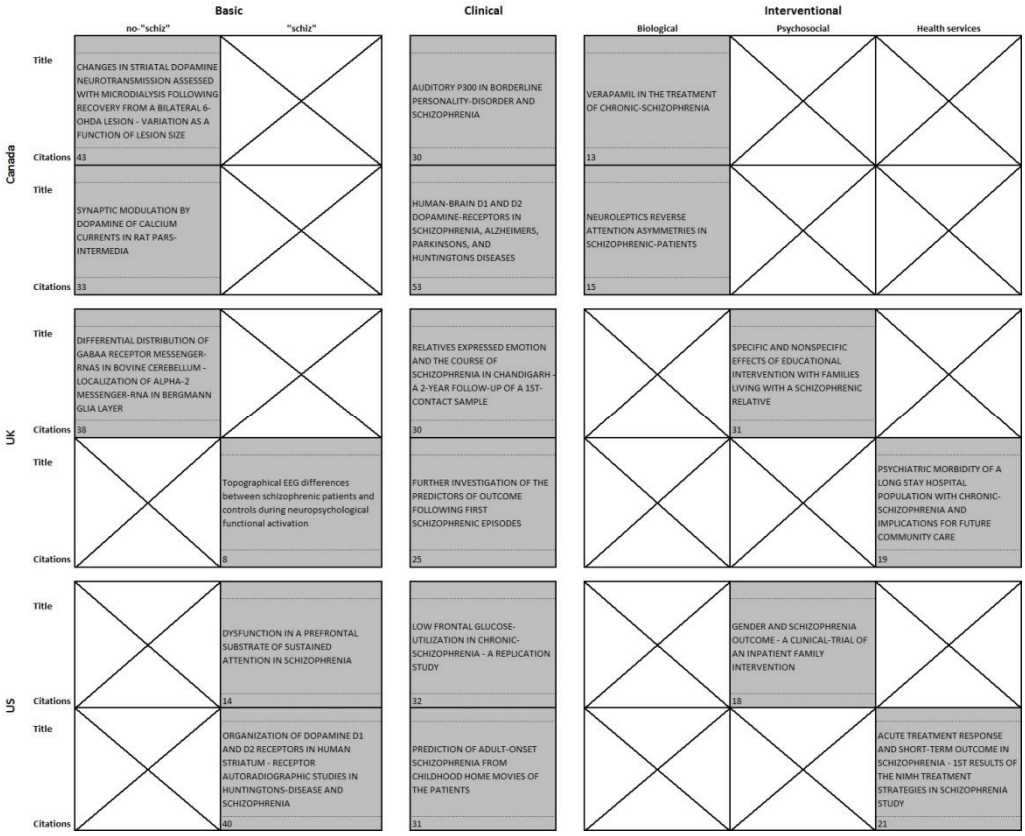


Figure A-2 Case study selection by country and research type

Having selected our full set of 18 case studies, we were able to contact the researchers involved to set up interviews and begin the process of developing the case studies.

Appendix B: Interview protocol for case study interviews

Introduction

Mental Health Retrosight is a multinational study that will evaluate the translation from basic or early clinical research into clinical application and practice. RAND Europe is leading the project in collaboration with a number of partner organisations.

RAND Europe is a non-profit policy research organisation, with an interest in research policy.

As part of this project, we are building a series of case studies around research that happened in the early to mid-1980s. A separate stream of work will track recent advances in mental health care backwards to identify the original research upon which the advance was based.

RAND Europe works in a number of policy areas. This work is being carried out within our Innovation and Technology Policy programme, which has a focus on research impact and policy. Our previous work in this area includes: payback studies of research impact on cardiovascular research, arthritis research and social science research. We also carried out a study for the MRC, Academy of Medical Sciences and Wellcome Trust to estimate the economic benefit of investing in medical research.

Using a bibliometric approach to identifying the most highly cited papers published in the late 1980s, we picked the paper you published in [BLANK]. We'd like to talk to you about how that research came about, how it fitted into the other research you were doing at the time and how it developed.

For this project, we are looking at both how the findings produced from the 'research cloud' were developed and translated; and, also, how the research undertaken as part of the 'cloud' developed the careers of researchers.

You should also emphasise 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.

You should also collect the signed consent forms that you sent to the PI before the interview.

Introductory questions

To begin, talk briefly about their current work and how it relates to what they were doing at the time.

1. Can you tell us a bit about what you were doing at the time?
2. Where you were in your career, etc.?
3. Can you give us some background to this paper?
4. Why do you think this paper was seen as important?
5. Why was this paper so highly cited?

STAGE 0: Opportunity Identification / Research Needs Assessment

Defining the cloud

1. What was the original impetus for the work?
   (Solely scientific curiosity? The desire to fill certain gaps in knowledge? Targeting of a particular disease state? Your own clinical experience?)
2. Did you see potential benefits for mental health at the time?

We want to try and define a block/unit/cloud of research which we can follow forward from. Can we talk about how the work for this paper fitted into the wider body of work you were doing at the time? And how that cloud of research might be delineated?

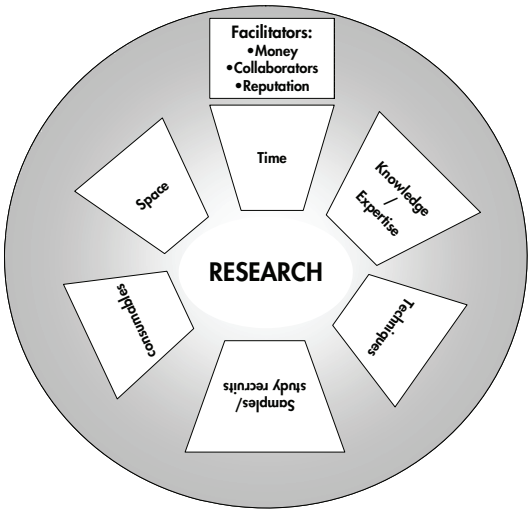
(Show paper list)

Influences on the cloud?

3. What other ideas were you pursuing at the time, how did they relate to this work?
4. How did this 'cloud' of research originate?
5. Who influenced your decision to work in this area?
6. Was it a continuation of previous work?
7. Who influenced this 'cloud'?
8. How far was your identification of the research topic influenced by:
   - Research you had done before? Funded by whom?
   - The research of others? If so, how did you hear about this research?
9. 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?
10. Did institutional conditions such as lab space, equipment or availability of researchers affect the research proposal?

STAGE 1: Inputs to Research and Project Specification and Selection

11. How was the cloud supported?
    - Ask for CV with grants etc.
    - What were the different forms of support and why was each important?
    - Did you make any unsuccessful applications for funding? Did you make any resubmissions?
    - Did any of the peer-review or applications processes affect the design or direction of the work?
    - Was there soft or core funding?
    - Did you have to compete for funding?
12. What was the institutional setting (hospital, university, research institute) for the research?
13. Who were the main researchers involved in the project?
14. What was their level of research experience and seniority at that time?
15. Had they previously worked in this research area?
16. Which of the following inputs were important (provide a copy of the diagram and discuss around the topics).



STAGE 2: Processes

17. Did the methods proposed prove to be appropriate? Which avenues of research were successful and which weren't?
18. Was there any interaction with potential users of the research *during* the research processes?
19. How much freedom did you or the research group have to pursue different lines of enquiry / deviate from the original proposal? How important was this flexibility in achieving the final results?
20. Did the research require new techniques / new expertise / new approaches to the subject?

21. How would you describe your role in the research process?
22. What was the role of collaborators in the research process (both academic and industrial)?
23. Who else was working in the area?

STAGE 3: Primary Outputs

24. Which publications do you think were most important from this research and why?
25. Did this work have any impact on the agenda for your subsequent research?
26. Did this research make any impact on the career of any of the research team? For example:
    - contribute to research training in terms of research degrees or the gaining of additional skills;
    - enable them to establish themselves in the field;
    - help the lead researcher to build a team of researchers.
27. 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?
28. Did the research spawn a new area of investigation or make a major impact on the approach used in subsequent research?
29. If the research was clinical, were any basic researchers also involved? If so, did this influence their attitude to clinical research?
30. 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?
31. 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*.
32. Have you had any impact outside the field of research you are working in?

INTERFACE B: Dissemination

33. 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?
34. Did you use specially designed dissemination approaches to particular audiences, for example policy briefs for policymakers? What were the most effective mechanisms for this?
35. What was the role of your networks in dissemination?
36. Did you receive support from funders/employers for dissemination? What form did this take?

STAGE 4: Secondary Outputs

37. 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?
38. Do you know how far the research directly influenced the formulation of any policy, or the realisation that a policy was needed?
39. Has any subsequent research by you or others that built on this 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?
40. Did the research from your project lead to any patent/licences? Was it taken up by industry? Has it contributed to any commercial products?
41. 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?
42. What **barriers** were there to the research having an impact / being translated?
43. What factors **facilitated** the research having an impact / being translated?
44. Has your research had an impact on teaching for clinicians?
45. Has any advisory role to government, hospitals or industry led to an impact from your research? If so, how did this come about?

Mapping exercise

Show the PI the bibliometric map of their paper and discuss the networks of co-cited papers that the analysis shows. Is there anything that surprises the PI about the map? Are there any papers that are being negatively cited?

STAGE 5: Applications

46. Have the findings from the research influenced practitioners directly through their reading the articles or hearing a presentation about the research?
47. 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?
48. Has any impact been local, regional, national or international?
49. 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?
50. Do you expect any greater take-up of the findings in the future? If so, where?
51. 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

52. 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 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?

53. Did engagement with the public/patient groups lead to changes in the researchers' perceptions of these groups?
54. 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

55. If the research has made an 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?
56. 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?
57. 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?
58. 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

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

Appendix C: Codebook

Code	Description	Unit of Analysis	Categories	Notes
Researcher characteristics				
Career	Initial career stage of researchers when working in cloud	Researcher	Early (i.e. post doc.), mid (i.e. between post and full professor), late (i.e. full professor), other	Code each researcher the first time they are mentioned; i.e. if a cloud has 3 main researchers, will each researcher be analysed on these characteristics?
Qualification	Highest degree/training	Researcher	PhD, MD, PhD or MD, other including other training experiences (please specify)	
Skills	Particular skills that the researcher contributes	Researcher	Tag & comment	
Motivation	Researcher motivation	Researcher	Curiosity focused / knowledge driven, inspired by interaction with patients, patient oriented, evidence of strategic thinking – i.e. being able to articulate the pathway to impact	Either individual or group; please note not mutually exclusive
Current role	Current role of researchers working in cloud	Researcher	Academic faculty, academic administration, retired faculty, government scientist, industry scientist, other (please capture)	

Initial role	Initial role of researchers working in cloud	Researcher	Academic faculty, academic administration, retired faculty, government scientist, industry scientist, other (please capture)	
Interest	Wider interests of researchers working in cloud	Researcher	Interests in other field(s) or areas (e.g. policy) (please capture field), research is focused on single topic, other (please capture)	
Record	Track record in field of researchers working in cloud	Researcher	New to field: i.e. no experience; less than 5 years in field; 5–10 years in field; long track record: i.e. more than 10 years in field	
Network	Networks	Researcher	Evidence of strong research links within an institution, evidence of research links across institutions, evidence of links with industry, evidence of links with policymakers, evidence of links with patients and families, evidence of other networks (please capture)	
Success	Successful researchers, unsuccessful research cloud	Researcher	Tag & comment	
Pathway	Researchers working in cloud involved in more than one stage of translation pathway	Researcher	Tag & comment	i.e. basic, clinical, implementation
Abroad	Researcher working in cloud has worked abroad	Researcher	Tag & comment	
Funding characteristics				
Review	Review of research by funder	Grant(s)	Review comments helpful, review comments unhelpful, review comments neutral, initial rejection, other (please capture)	
Process	Type of funding process	Cloud	Response mode (investigator initiated), specific/target funding opportunity, other (please capture)	

Type	Types of research funding	Cloud	Project / grant in aid (i.e. up to 4 years), programme (i.e. 4 years +), fellowships, Institutional (i.e. research centre), no formal funding, other (please capture)	
Multisite	Part of a multisite study	Cloud	Tag & comment	
Sources	Sources of research funding	Cloud	Government, foundation, industry, other (please capture)	
Size	Length/value of research funding	Cloud	Please capture numerically, or unknown category	
Cuts	Suggestions that funding was cut by review committee	Grant	Tag & comment	
Resources	Comment on level of resourcing of research	Cloud	Tag & comment	e.g. was under-resourced, had plenty of funding
Leverage	Suggestions that initial funding leveraged out subsequent funding	Cloud	Tag & comment	
Views	Researcher comments on funding sources	Cloud	Tag & comment	e.g. about a specific source being particularly important
Competition	Researcher comments on degree of competition at time for funding	Cloud	Comment that highly competitive, competitive, or not competitive	
Design	External input to design of research	Cloud	Tag & comment	e.g. input from funder, from patient groups
Research characteristics (some of these replicate researcher characteristics but these are specific to the cloud)				
Combine	Suggestions that basic and clinical approaches are combined	Cloud	In approach, in researchers (i.e. in individual researcher, in team, in location), other	
Engage	Engagement in research process	Cloud	Scientists, clinical researchers, practitioners, patients, families, policymakers, media, other	

Hypothesis	Suggestions about the nature of research / scientific hypothesis	Cloud	Safe, fashionable, controversial, contrary to established view, high risk, other	Please note that categories are not mutually exclusive: i.e. code more than one if necessary
Serendipity	Serendipitous event	Cloud	Tag & comment	Please note that this may also refer to something not related to the research: e.g. a chance meeting
Collaboration	Collaboration	Cloud	International, national, local: i.e. state/provincial, infrastructure, materials: e.g. brain bank, other	
Wider	Research part of a wider coordinated and directed programme	Cloud	Tag & comment	
Interdisciplinary	Suggestion of interdisciplinary research	Cloud	Tag & comment	
Application	Suggestions of application of new method in field	Cloud	Tag & comment	Please capture timescales: e.g. if method was overtaken by new technologies
Subject	Subject of research	Cloud	Cell, animal tissue, animal, human tissue, human, other	
Changed	Suggestion that research plans changed from original ideas	Grant	Tag & comment	
Unplanned	Suggestion that research did not work out as planned – for better or worse	Cloud	Better, worse, other	
Champion	Senior champion	Cloud	Research, implementation, other	
Uptake	Researcher comments on reasons for uptake/development (or lack of): e.g. superseded by something else	Cloud	Uptake or development, no uptake or development, other	
Institutional characteristics				

Intensity	Research intensity of institution	Cloud	Top 10%, top 25%, top 50% of research funding in country, other	May wish to tag & comment and estimate – the actual distributions can be worked out later
Setting	Setting of research	Cloud	University, hospital, institute, clinical, other (please capture)	
Culture	Comments on context / culture of research	Cloud	e.g. culture of lab, institutional reputation, access to infrastructure, location of research (e.g. clinical setting)	
Infrastructure	Available infrastructure	Cloud	Biostatistics / design support, grant writing, lab space, specialised equipment, other (please capture)	
Focused	Mission of research institute is focused or not focused on research interest of the cloud	Cloud	Focused on same interest, not focused on same interest, other	
System characteristics				
Regulation	Effects of regulation	Cloud	Human subject protection issue, animal subject protection issue, other	
Systems	Influence of other systems	Cloud	Healthcare, media/public perception, political, economic, social, policy, other	
Historical	Changes in system mean that it no longer operates in this way	Cloud	Tag & comment	
National	Cultural characteristics specific to each country	Cloud	Tag & comment	
Shift	Paradigm-like shift in understanding in field	Cloud	Tag & comment	e.g. a shift towards a biological explanation of schizophrenia
Other				
Interesting	Interesting and notable	Any	Tag & comment	
Quotes	Please tag all quotes	Any	Tag & comment	

Appendix D: The rating workshop

Once initial ratings had been given, a workshop brought together the panel to discuss the case studies. The aim of this session was to reduce any differences in understanding of the research described in the case studies, but preserve any differences that remained in raters' impact scores that remained after discussion.

The workshop took place over two days, 13–14 September 2012, and eight of the nine panel members were able to attend. It was not possible to discuss every single rating for each case study in the time available, so it was necessary to prioritise the discussion. This was done on the basis of two criteria: level of disagreement between scorers, and suggestion from scorers.

The discussion was structured by payback category, and started with an overall review of the scoring process and any clarifications required around the definition of the category and the different approaches taken to the task by panel members. Then, the case studies which had the highest level of disagreement within the category were discussed, along with those case studies suggested by panel members as requiring further discussion. The level of disagreement between scorers was assessed through the average deviation from the median score of the scores assigned to a particular case study in each category (ADM). These data are presented in Figure D-1, with the case studies listed in order of ascending ADM in each category. We considered two possible mechanisms for selecting case studies for discussion. The first was to take a cut-off value of 1.1 for all categories and discuss all case studies with a higher level of disagreement than this (shown in blue in Figure D-1). However, we decided instead to go with an approach whereby we attempted to discuss a roughly even number of case studies under each category, setting the cut-off point as necessary to best achieve this in each category (shown in red in Figure D-1). In addition, we discussed case studies suggested by the panel (marked in green in Figure D-1). We also retained some flexibility to discuss issues as they arose and to adjust our approach depending on the progress of the workshop.

KP	RTCB	IPPD	HHSB	BEB	OI
0.5 M	0.4 J	0.4 A	0.3 M	0.3 N	0.4 M
0.6 H	0.5 M	0.4 M	0.4 A	0.4 B	0.5 N
0.8 E	0.6 N	0.4 N	0.4 B	0.4 M	0.6 G
0.9 J	0.8 P	0.6 J	0.5 K	0.6 A	0.7 A
1.0 D	0.9 G	0.9 B	0.5 N	0.6 G	0.7 P
1.0 P	0.9 L	0.9 F	0.9 G	0.6 K	0.8 D
1.1 C	1.1 A	0.9 G	0.9 J	0.6 Q	0.8 K
1.1 G	1.1 D	0.9 P	0.9 P	0.8 C	0.8 J
1.1 I	1.1 I	1.3 K	1.0 H	0.8 D	0.9 Q
1.1 K	1.3 B	1.4 O	1.1 F	0.8 H	1.0 B
1.1 N	1.3 E	1.5 D	1.3 C	0.8 O	1.0 F
1.3 A	1.3 F	1.6 H	1.4 D	1.0 F	1.1 L
1.4 F	1.3 K	1.6 Q	1.4 E	1.1 E	1.1 C
1.5 L	1.4 H	1.8 E	1.4 O	1.1 J	1.1 E
1.5 Q	1.4 Q	1.8 I	1.6 Q	1.4 P	1.2 H
1.8 O	1.8 O	1.8 L	1.9 L	1.5 L	1.3 O
2.3 B	2.0 C	1.9 C	2.1 R	2.0 R	1.8 R
2.3 R	2.4 R	2.3 R	2.3 I	2.3 I	2.0 I

Figure D-1 Case studies listed by Average Deviation from the Mean in each category, with selection criteria overlaid (case studies anonymised)

To support the discussion, scorers were given sheets which showed the distribution of scores for each case study in each category, with their own score indicated by a blue line within the distribution. Others' scores were not shared. This was used as a prompt for discussion.

Following the discussion in each category, scorers were given the opportunity to rescore if desired on the basis of the discussion, in confidence. These revised scores were then collected by the team at the end of the session.

In the overall impact category discussion, we also asked scorers to describe how they reached their judgement of overall impact. We had not provided specific guidance on how to do this, and it was interesting to note that different scorers had taken very different approaches. Some scorers had just averaged scores across categories. Others had weighted more downstream impacts more heavily (e.g. broader economic benefits or health sector benefits), and some had weighted knowledge production more heavily. This lack of consensus about the relative weighting of different elements within the overall impact score led us to decide that the overall score would not be used in the later analysis.

Scoring results: level of consensus and correlation between categories

The scores generated after the workshop were analysed to assess first the level of agreement between the scorers, and whether this had changed following the workshop, to establish whether there was a consensus on the level of impact of each case study in each category; and secondly to look at the level of correlation between the different categories, to inform our approach to analysis in the subsequent stages of the project.

The correlation between the final set of scores produced following the workshop is shown in Table D-1. The level of correlation between all scorers is strong and statistically significant at the 1% level. Table D-2 provides descriptive statistics regarding the scores of each scorer. The data included here are the final set of scores for the seven scorers who attended the workshop, plus the scores for the one scorer who provided scores but did not attend the workshop (scorer 6). We find that the overall correlation between the scorers is sufficient to indicate consensus and that the scores from the one scorer who did not attend the workshop do not appear to be substantially different from those of the other scorers, so we have included them in the subsequent analysis.

Table D-1 Correlation between scorers after workshop.[8] The correlation was calculated using a two-tailed Spearman's Rho in SPSS. Those correlations which are statistically significant at the 1% level are marked with two stars, which is true for all correlations in this case.

	Scorer 1	Scorer 2	Scorer 3	Scorer 4	Scorer 5	Scorer 6	Scorer 7	Scorer 8
Scorer 1	1.000	0.857**	0.858**	0.710**	0.840**	0.862**	0.861**	0.815**
Scorer 2	0.857**	1.000	0.826**	0.796**	0.817**	0.804**	0.860**	0.846**
Scorer 3	0.858**	0.826**	1.000	0.771**	0.882**	0.847**	0.852**	0.828**
Scorer 4	0.710**	0.796**	0.771**	1.000	0.765**	0.697**	0.767**	0.762**
Scorer 5	0.840**	0.817**	0.882**	0.765**	1.000	0.876**	0.867**	0.811**
Scorer 6	0.862**	0.804**	0.847**	0.697**	0.876**	1.000	0.780**	0.821**
Scorer 7	0.861**	0.860**	0.852**	0.767**	0.867**	0.780**	1.000	0.803**
Scorer 8	0.815**	0.846**	0.828**	0.762**	0.811**	0.821**	0.803**	1.000

Table D-2 Descriptive statistics for each scorer (after workshop)[9] – '0.25' – lower quartile, '0.75' – upper quartile, IQR – inter quartile range

	Scorer 1	Scorer 2	Scorer 3	Scorer 4	Scorer 5	Scorer 6	Scorer 7	Scorer 8
Median	5	4	3.5	5	4	3	4	3.5
Mean	4.3	4.1	3.4	5.2	4.0	3.4	4.1	3.7
0.25	1	2	0	3	1	0	2	1
0.75	7	6	5	7	7	6	6	6
IQR	6	4	5	4	6	6	4	5

[8] This includes the rescores for those who attended the workshop, and the unaltered scores for the scorer who submitted scores but was unable to attend the workshop.

[9] Where available. scores for the eighth scorer who did not attend the workshop also included.

To illustrate the increase in agreement between scorers resulting from the workshop discussions, Table D-3 provides the medians for each category for each scorer both before and after the workshop. The mean of the ADM values across all case studies and categories fell as a result of the workshop, from 1.1 to 0.9. This is calculated by first generating an average deviation from the median (ADM) score for each case study in each impact category. The average ADM for each category is then calculated, and finally these category scores are averaged to give a measure of the overall level of disagreement between scorers across all categories and case studies.[10] The level of agreement achieved is similar to that found in the previous Cardiovascular Retrosight study.

Table D-3 Changes in median score as a result of workshop for all scorers providing scores both before and after workshop

		KP	RTCB	IPPD	HHSB	BEB	Overall	Median across all categories
Scorer 1	**Before workshop**	6	5	4.5	1	0	4	4
	After workshop	6	5	4	2	0.5	5.5	5
Scorer 2	**Before workshop**	5	4	4	2	1.5	4	4
	After workshop	5.5	4	4	2	1.5	4	4
Scorer 3	**Before workshop**	4.5	4	1	0	0	4	3
	After workshop	4.5	4.5	2.5	0	0	4	3.5
Scorer 4	**Before workshop**	6	7	5.5	5	2.5	5.5	6
	After workshop	6	7	5	5	2.5	5	5
Scorer 5	**Before workshop**	6	6	4	1	0	4	4
	After workshop	6	6	4	1	0	4.5	4
Scorer 6	**Before workshop**	5	5	4	2.5	2	4.5	4
	After workshop	na	na	na	na	na	na	na
Scorer 7	**Before workshop**	6	5.5	3.5	0	1	3.5	3
	After workshop	5.5	5	4	1	1	4.5	3.5

Assigning case studies to high- and low-impact groups

Having established that there was reasonable consensus between scorers, the next step was to decide on the most appropriate way to conduct the subsequent analysis. First, we looked at the level of correlation between the different impact categories used, as shown in Table D-4. We found that there was a statistically significant level of correlation between most categories at the 5% level at least, but typically the correlation is strongest with adjacent categories, and in all cases the correlation with adjacent categories is statistically significant. For example, research targeting and capacity building (RTCB) is correlated most strongly with knowledge production (KP) and informing policy and product development (IPPD),

[10] Note that pre-workshop not all scores followed the scoring protocol suggested and in some cases did not use the full range of scores. For this calculation, pre-workshop scores used were actual scores provided (not normalised), with the exceptions of: one score of 10 changed to 9; overall impact, for which normalised scores were used.

and less well correlated with health and health sector benefit (HHSB) and broader economic benefit (BEB). Similar relationships are found throughout. In addition, all categories have a statistically significant correlation with overall impact (OI) at the 1% level. However, we know from discussion with our scorers that the way in which the OI score was generated differed significantly between scorers, although it always involved some form of averaging across the category scores, so we choose not to use this score in our further analysis.

Table D-4 Level of correlation between categories using a two-tailed Spearman's Rho. Correlations significant at the 5% level are marked with one star; correlations significant at the 1% level are marked with two stars.

	KP	RTCB	IPPD	HHSB	BEB	Overall
KP	1.000	0.836**	0.499*	0.282	0.429	0.643**
RTCB	0.836**	1.000	0.486*	0.364	0.505*	0.593**
IPPD	0.499*	0.486*	1.000	0.803**	0.818**	0.858**
HHSB	0.282	0.364	0.803**	1.000	0.952**	0.778**
BEB	0.429	0.505*	0.818**	0.952**	1.000	0.859**
Overall	0.643**	0.593**	0.858**	0.778**	0.859**	1.000

Since all categories of impact are correlated to some extent and there is no obvious cut-off point or clear way to separate impact categories, we have chosen to analyse the impact in each category individually. To do this, we next identified a group of high- and low-impact case studies in each category. We did this by ranking each case study in each category individually, and then dividing each category into thirds, or as close as possible, keeping tied rankings together. This generated a high-impact, low-impact and mid-impact group of case studies for each impact category. There are between five and seven case studies in each group. In the analysis, we focused on the high- and low-impact groups, comparing factors that are present or absent in each of these groups by category.