

WORKING P A P E R

Possible approaches for evaluating Arthritis Research Campaign grants

An options paper

STIJN HOORENS AND STEVEN WOODING

WR-662-ARC

February 2009

Prepared for the Arthritis Research Campaign

This product is part of the RAND Europe working paper series. RAND working papers are intended to share researchers' latest findings and to solicit additional peer review. This paper has been peer reviewed but not edited. Unless otherwise indicated, working papers can be quoted and cited without permission of the author, provided the source is clearly referred to as a working paper. RAND's publications do not necessarily reflect the opinions of its research clients and sponsors. RAND® is a registered trademark.

Contents

Preface	4
CHAPTER 1 Introduction	5
1.1 Identifying outcomes of research	6
1.2 Current practice in evaluating outcomes	7
1.3 Aims and objectives.....	8
1.4 The ideal system.....	9
CHAPTER 2 Issues in research evaluation	10
2.1 General issues in research evaluation	10
2.2 Attribution.....	10
CHAPTER 3 Tools for evaluating research	14
3.1 Tick list and menu	14
3.2 Exemplar scales	15
3.3 Calibrators	16
3.4 Pathways	17
3.5 Matrices	19
3.6 Short-listing promising projects	20
3.7 Searchable repository.....	21
CHAPTER 4 When to collect information	22
APPENDICES	25
Appendix A: Key Informant Interviews	26
Appendix B: Draft tick lists	27
Appendix C: Example exemplar scales	29
1. Knowledge Production	29
2. Research Targeting and Capacity Building	30
3. Informing Policy making and Product Development	30
4. Health and Health Sector Benefits.....	32
5. Broader Economic Benefits	33

Preface

The Arthritis Research Campaign (**arc**) has asked RAND Europe to help it develop a system to monitor the outputs and outcomes of the research it funds. To assist with this process an Advisory Group of **arc** funded researchers has been set up to provide additional scientific expertise to the project team. This report was produced as a background document for the first Advisory Group workshop. The report outlines the project; introduces the important concepts in research evaluation, and describes some potential research evaluation tools that were explored in the workshop.

This report will be of interest to other research funding agencies and evaluators who are keen to measure the impacts of research.

The research was led by RAND Europe, with additional input from the Health Economics Research Group at Brunel University and the newly appointed Medical Director of **arc**, Professor Alan Silman. 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, multidisciplinary analysis. This report has been peer-reviewed in accordance with RAND's quality assurance standards.

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

Steven Wooding
Team Leader, Science and Technology
RAND Europe
Westbrook Centre
Milton Road
Cambridge CB4 1YG
United Kingdom
Tel. +44 (0)1223 353329

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

The Arthritis Research Campaign (**arc**) would like a better understanding of how the research it supports leads to improvements for arthritis sufferers. An overview of the outputs and outcomes, and likely outcomes, of **arc** research grants, is required in order to develop this understanding. Such an overview should include direct outputs, such as publications, but must also wider outcomes, such as influence on clinical practice; input into drug development; and benefits to patients. Such a system will never be perfect at capturing and assigning impact, but it should provide an improved evidence base for fundraising, lobbying and strategic decision making. Clearly, this is a difficult problem and one that has not, to our knowledge, been successfully solved by any other research funder.

From our initial work it is clear that such a system would be composed of a combination of different evaluation tools and RAND is asking the Advisory Group to help with a preliminary evaluation of some promising tools. RAND hopes the Advisory Group then continue to assist with the further development of those tools towards a practical evaluation system.

This document sets out the background to this project and outlines the work completed to date. This work has been informed by RAND's previous work in research evaluation, assessment and strategy - including work done for **arc**; the MRC; the Department of Health and the National Audit Office. More specifically this work draws on discussions with **arc** and a small group of key individuals identified by **arc** representing many of the key constituencies in arthritis research (see Appendix A). This was backed up with a scan of the literature and interviews with other research funders to look for helpful examples.

The introduction provides the background to the project; outlines current **arc** monitoring procedures and evaluations; and describes the rationale of developing these further; it then discusses the aims of improved monitoring and what an ideal system might look like, before touching on three of the overarching issues of research evaluation. Based on this document, RAND will work with the Advisory Group to select two or three promising tools. These will then be further optimised and tested in collaboration with the Advisory Group and developed to produce a monitoring system.

1.1 Identifying outcomes of research

Health research funding bodies are under increasing pressure to demonstrate the outcomes, or benefits, of the research that they fund.¹ Traditional peer review of research focused on the outputs in terms of journal articles, the training of future researchers and the development of careers. While these are still seen as important, it has been widely recognised that this is a very narrow definition of the paybacks from scientific research and that it is important to look far more broadly at the benefits flowing from scientific research. One effective method of doing this, which will be used in this project, has been to use a broader multidimensional categorisation of the benefits from health research, such as was used in the work examining the long-term outcomes of **arc** research².

We have used the HERG/RAND payback framework as a starting point for identifying the outcomes of arthritis research.³ The Payback Framework uses 5 categories to classify the impacts of research. These categories reflect the objectives of the research programme and can include quantitative measures of impacts (for example, the percentage of PIs who claim that their research has made some impact on policy) and qualitative descriptions of impacts (for example a case study description of how the research informed a policy debate). The categories aim to cover the full range of research impacts starting with knowledge production and running through to economic benefits. In outline the categories are:

- **Knowledge Production:** the initial output of research. This knowledge is often codified in papers and presentations. Knowledge production can be assessed by looking at those outputs, and assessing them either qualitatively, through peer review, or quantitatively using techniques such as bibliometrics.
- **Research Targeting and Capacity Building:** the way in which the research raises or focuses new research questions and the contributions it makes to providing the skills or people to undertake future research.

¹ See for example:

Buxton M, Hanney S *et al.* Assessing benefits from Department of Health and National Health Service Research & Development. *Public Money & Management* 2000; 20:29-34.

Grant J, Cottrell R *et al.* Evaluating "payback" on biomedical research from papers cited in clinical guidelines: applied bibliometric study. *BMJ* 2000; 320:1107-1111.

Croxson B, Hanney S, Buxton M. Routine monitoring of performance: what makes health research and development different? *J Health Serv Res Policy* 2001; 6:226-232.

Smith R. Measuring the social impact of research. *BMJ* 2001; 323:528.

Weel C van. Biomedical science matters for people—so its impact should be better assessed. *Lancet* 2002; 360:1034-1035

² Wooding S, Hanney S, Buxton M and Grant J (2004). *The Returns from Arthritis Research. Volume 1: Approach Analysis and Recommendations.* RAND Document MG-251-ARC. Santa Monica: The RAND Corporation. Available from: <http://www.rand.org/publications/MG/MG251/>

³ Hanney SR, Grant J, Wooding S, Buxton MJ. Proposed methods for reviewing the outcomes of health research: the impact of funding by the UK's 'Arthritis Research Campaign'. *Health Research Policy and Systems.* 2004 Jul 23;2(1):4.

For full text see: <http://www.pubmedcentral.gov/picrender.fcgi?artid=503400&blobtype=pdf>

- **Informing Policy and Product Development:** the ways in which the research has influenced policy (in the broadest sense) and/or how it has contributed to the development of new pharmaceuticals or devices that could improve treatment.
- **Health and Health Sector Benefits:** the contribution of the research to improvements in health and/or increases in the effectiveness and equity of the health system.
- **Broader Economic Benefits:** the contributions such as exports that can result from the commercial exploitation of the research and the value of the production gained from having a healthy workforce.

1.2 Current practice in evaluating outcomes

arc's current practice

arc already has well developed systems for collecting information about the initial scientific outcomes of grants, systems which are ahead of many other research funders. **arc** routinely collects information in three ways: Firstly, **arc** has a long history of collecting end of year reports for its grants and publishing them in book form - providing a (pre-internet) means to locate researchers with specific interests. Secondly, **arc** requires end of grant reports detailing what has been achieved; and finally, **arc** requires information on the outcomes of previous grants to be provided on new applications for funding. However, this information has never been analysed across all (or most) of the research that **arc** funds and although track record may be taken into account in the award an individual grant, it has not been used to inform strategic decision making.

In addition to this regular information collection, **arc** has carried out a number of ad hoc evaluations including the work with RAND on the long term outcomes of research, internal evaluations of fellowships career trajectories and bibliometric analysis of **arc** funded publications. Although valuable, these studies have only looked at a small sample of **arc** funded research, or have examined only a small range of outputs (publications, for instance).

Current practice of other funders

Over recent years medical research funders have become far more concerned about evaluation and examining the outcomes, as well as the outputs, of medical research. However, we were not aware of any other charitable foundations or funding agencies that carry out systematic review of the wider impacts of their research activities. To ensure we were not reinventing the wheel we carried out interviews with six foundations who we considered were most likely to have interesting evaluation procedures, based on our knowledge of selection procedures and ethos.⁴

Four of the six are interested in developing methods to systematically assess the wider impacts of their research; however, none of these initiatives have proceeded very far. The

⁴ Funders interviewed were: Breakthrough Breast Cancer; The Canadian Health Services Research Foundation; The Canadian Institutes of Health Research; The Joseph Rountree Foundation; The National Health and Medical Research Council of Australia and The Parkinson's Disease Society.

National Health and Medical Research Council of Australia (NHMRC) has the most developed initiative and this aims to develop a metric to evaluate the impact track record of researchers as one input to the peer review process. This metric aims to take into account a wide range of research impacts from knowledge production through to economic benefit and uses similar concepts to those we introduce in Chapter 3. All of the foundations required end of grant reports, some also asking for intermediate reports, but these were restricted to the research results and the immediate publications. And all have carried out ad hoc evaluations to look at particular aspects of their work, or particular collections of grants.

Drivers for change

There are a variety of drivers for improving the understanding of the impacts of **arc** funded research, which can be separated into three main categories. From a fund raising perspective it is important to be able to show how research funding is linked to improvements in the treatment and prevention of arthritis. There is also increasing pressure from organisations such as the Charity Commission to demonstrate how different types of research contribute to the stated mission of being 'Committed to curing arthritis'. Finally, it is important for **arc** to have an evidence base on which to make strategic decisions about the most appropriate areas and types of funding. A related issue is the concern that opportunities to develop the knowledge or expertise produced in **arc** funded research towards patient benefit may be being missed. There should be system that actively identifies promising opportunities for developed.

1.3 Aims and objectives

The system would have three principal (and interrelated) aims:

1. To provide an on-going assessment of the wider impact **arc** funded research. The idea is that this would be a forward-looking assessment, starting with current research and following it forwards in time, rather than looking at past research.
2. To attempt an estimate of the future wider impacts of the research to provide an impression of the benefits that might accrue in the future.
3. To provide mechanisms to improve the chances of new knowledge or expertise being developed towards patient benefit.

Each of these aims has two building blocks:

- Firstly, a mechanism to collect information on the research and its outcomes.
- Secondly, a mechanism to aggregate or score this information to provide a means to: summarise; compare the outcomes of collections of grants; or to identify grants with particular potential for development.

1.4 The ideal system

Our background work suggests a series of characteristics for an ideal system; however, it is also clear that it may not be possible to achieve all of these characteristics in one system and that compromises may need to be made. An ideal system would have the following characteristics:

Capturing the full range of benefits

From our previous work on research assessment it is clear that research produces many benefits and impacts beyond publications and research qualifications, the system should seek to capture as many of these as is feasible.

Aggregation

The system should allow impacts of many grants to be aggregated together to provide an impression of the overall impact of a group, or of all, grants. But it should also allow impacts of very different types to be kept apart – for example the production of knowledge and influence on health policy. This would allow the different strengths of different types of research to be explored.

Valuation

The system should provide a way to consider the differing value of different types of impacts, ie a method of reducing a range of impacts to a common currency. However, in doing so it should be sensitive to the fact that different groups, such as researchers, physicians or patients, might perceive and value impacts differently.

Low Burden

While any system is bound to require some time and resources from both researchers and administration, this burden should be in proportion to the benefit that the system provides. It is important only to collect information that can be used and not collect information because the opportunity exists.

Wide applicability

The system should be widely applicable to all forms of **arc** research, although some variation across the different areas of research may be necessary

Fairness

The system should capture information in a way that is fair and consequently allows comparisons of groups of research grants or types of research.

Timeliness

The speed with which the system can provide information will always be a trade off between the requirement for speed to support decision making, and allowing time for the outcomes of research to develop. Because of this a monitoring system needs to try and provide early indicators of later impact where this is possible.

Although there are many issues of detail that stem from the evaluation tools selected, there are also a few general issues that concern all forms of research evaluation. As these general issues affect all further discussion in this chapter we outline them briefly and discuss their relevance to this project.

2.1 **General issues in research evaluation**

Accuracy

As mentioned earlier in this paper an evaluation system can never be perfectly accurate, but the accuracy needs to be sufficient for the purposes to which the information will be put. This is: demonstrating the effectiveness of **arc** funded research, justifying the charitable nature of **arc** funding and informing strategic decision making. It is crucial to bear in mind that the information will be used in an aggregated form and hence small inaccuracies or omissions (provided they are not systematic or biased) are acceptable.

Prediction and timescale

One of the complications of research evaluation is that many of the impacts of research occur long after the research has taken place, this suggests that the system may need to be able to track research over the long-term to build up a full picture of its impact. There has also been much discussion about how accurately the impacts of research findings can be predicted. There are examples both where impact was easy to predict and others where prediction would have been impossible. However, we suggest that if predicting future impact is seen to be an important function of such a system, then the most effective way to address this question is to test whether predictions of researchers (or others) are accurate or not.

Attribution

Attribution is one of the thorniest issues in research assessment and is discussed in more detail in the next section.

2.2 **Attribution**

One of the recurring questions in research evaluation is attribution: how strong is the link between a particular research discovery and a particular change in health care, or between a particular grant and a particular scientific discovery? Attribution is both philosophically

and practically complex. However, as there is rarely a one-to-one correspondence between a scientific discovery and an advance in healthcare, we need a method of dealing with it.

Attribution can also be discussed in terms of contribution, the key difference being the direction of the causal link being explored. If we are looking at how a particular research grant contributed to a discovery, we would use the phrase ‘the grant’s contribution’; however, if we were looking at how much a particular discovery owed to a specific research grant we would talk instead of how much of the discovery could be attributed to the grant.

In this chapter we discuss some of the issues involved in dealing with attribution, highlight in what circumstances they are important and suggest a couple of promising approaches

The complexity of attribution

The complexity of dealing with attribution stems from the fact that most developments have many causes and each of these causes will also have come about due to a series of prior circumstances. Each of these pyramids of causation will stretch indefinitely into the past becoming increasingly diffuse as they so. In order to restrict this work to manageable dimensions when looking at the outputs and outcomes of a grant we suggest using the concept of the counterfactual – considering what would not have happened had the research not been funded. However, although this restricts the number of aspects to consider, it does not completely simplify the issue.

For the purposes of illustration we will start with a simple example and gradually add complexity to it. First off consider the example in Figure 1: a laboratory funded entirely by an **arc** programme grant, which carries out a clinical trial leading to a change in NICE guidelines on the management of back pain.

Although the chain of attribution seems straightforward, there are still two steps that need to be considered: the link between the funding and the research and the link between the research and the health outcome (in this case a NICE guideline).

The situation becomes more complicated if a second source of funding is involved (Example 2 in Figure 1), for example if the lead researcher is an MRC research fellow. However, as support will generally be in terms of money one approach would be to consider the relative contribution of the two funding streams in proportion to their size. However, this approach seems unreasonable where, a minor fraction of the support - possibly a particular piece of equipment - although small in financial terms, may be crucial in allowing the research to be carried out.

Looking at the link between research and outcomes is more complex still, but for similar reasons. In our example with one clinical trial leading to one guideline the attribution is relatively clear. However, there will generally be more than one piece of research that helps drive a change in policy or practice (Example 3 in Figure 1). In such a case each piece of research is likely to have contributed something to the change. We could attempt to provide a crude estimate of this contribution by classifying each piece of research as a major or minor contributor to the change. We could even go further and provide a subjective valuation of the contribution in percentage terms. However, it soon becomes clear that such numerical scales are problematic.

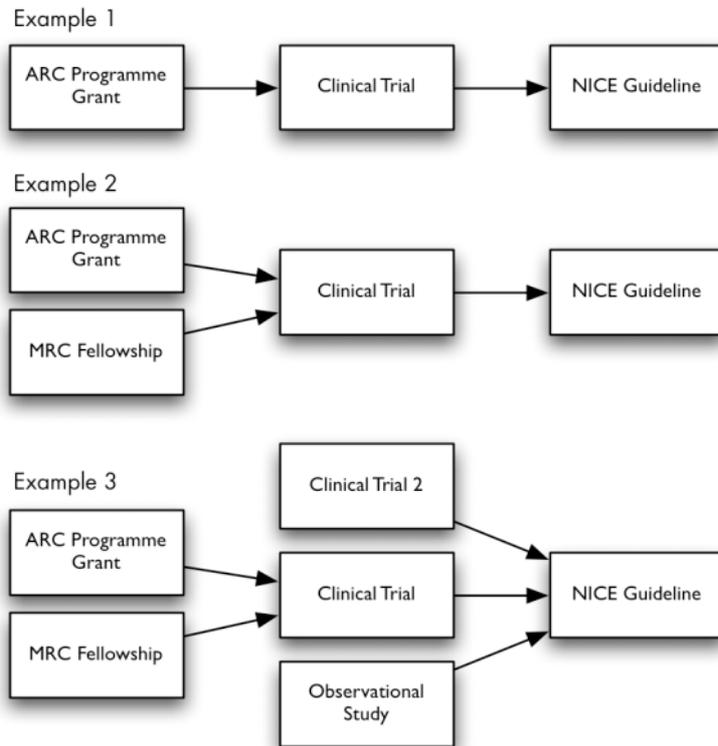


Figure 1: Examples of attribution

Consider the example where the two clinical trials would have been enough evidence on their own to cause the change, but the observational study provided evidence in a readily accessible form that greatly accelerated the change. How could this be handled using a numerical division of contribution? There might of course be other non-research factors that contributed to the change; however, for the time being we are considering only research factors.

An alternative way to consider attribution - rather than a fractionation approach - might be to consider how long a particular change, such as the rewriting of the guideline, might have taken if the research had not been available: would it have take another year of research accumulation or another five. However, although attractive, this option suffers from a number of disadvantages including a problem with highly competitive fields. If a number of research groups are competing in a field, the length of time the field would be delayed by the failure of any one of them to publish, is likely to be much less than the length of time a less competitive field would be delayed.

So far, we have discussed how attribution could be determined or quantified, the final issue is how to take it into account when aggregating outcomes or impacts of research. In order to allow comparisons between groups of research projects it is necessary to convert the impacts of research into a common currency, or a series of metrics each relating to a different type of outcome. When calculating the total impact of a piece of research it would seem sensible to multiply the value of its impacts by the degree to which the research contributed to each impact. In other words, the total value of the impact of a piece of research is the value of each of its individual impacts multiplied by the contribution that the research made to each impact.

It seems clear that for each piece of research the level of contribution should be used as a multiplicative factor when combining an assessment of the size of the impacts of the research. But how many minor contributions are equivalent to one major contribution? As discussed previously this is probably not a linear scale.

When does attribution matter?

The different aims of our exercise place different requirements on how we deal with attribution. In order to identify opportunities for further development, attribution can be ignored; we are only interested in good opportunities, not how they came about. From an accountability point of view, from both a donor and charity commission point of view, attribution is important, but a detailed analysis of it may not be. The key question here is, how **arc** research is contributing to improvements in healthcare, the exact strength of the link is less important. Attribution is most important when providing information to inform strategic decision making, in particular when comparing different groups of grants.

It is very important to bear in mind that for all of these uses the data will be used at an aggregated level. In this case each measurement does not need to be perfect, but we need to avoid a system that is systematically biased.

Possible ways forward

In considering these issues, we have come up with two promising ways forward, and we will be testing aspects of them throughout the workshop. In light of the discussion above, we think it is unlikely that we can come up with objective definitions for a scale for attribution, so we present two alternative ideas:

Use subjective threshold

In this case we would only include outcomes or impacts that reached a subjective threshold of attribution. For example, only report outcomes for which the research had been a significant contributor. This approach has the advantage of simplicity, but discards information about the level of attribution and other impacts to which the research made only a minor contribution.

Use broad subjective categories and rely on common understanding

In this case we would ask for a subjective evaluation of the strength of the attribution on a scale such as 'peripheral', 'minor', 'major', 'crucial'. This has the advantage of not discarding information, but the disadvantage of complicating aggregation and increasing the amount of information that we would need to collect.

One way to address the issue of aggregation would be to explore whether people attach similar relative numerical values to the terms used. This will be discussed as part of the workshop.

As explained above, a system for evaluating **arc** grants will have multiple aims: collecting the information on the variety of paybacks that have been achieved; quantifying those paybacks to allow for management decisions; and identifying research projects that have potential for further development. In previous projects we have shown that although scoring cannot differentiate between all projects, it can usefully summarise project outputs and scorers can consistently differentiate between bands of projects. There are a number of evaluation tools we could use and each addresses one or more of the aims. Below we describe a selection of these tools, along with some notes on their likely advantages and disadvantages. In any actual system we would be likely to use a combination of tools.

3.1 Tick list and menu

This tool would include a very detailed list, comprising a (nearly) comprehensive set of likely paybacks. Researchers would then be asked to indicate which of these paybacks their grant has produced, possibly along with a very brief commentary. For example: 6 peer-reviewed publications, 5 international conference presentations, 3 PhDs trained, 2 follow-up grants, etc. A preliminary tick list for **arc**-funded research is provided in Appendix B.

Tick lists are particularly useful for cataloguing outputs from the first two payback categories (Knowledge Production and Research Targeting and Capacity Building). While ‘citations in clinical guidelines’ or ‘diagnostic tests developed’ are possible items on the tick list for Informing Policy and Product Development, the outputs become more diffuse in the last three payback categories. We suspect it would be very difficult to capture all relevant items in these later categories.

This method could be combined with a ‘menu list’ quantifying tick list items to normalised scores, in order to allow aggregation of all outcomes and comparison across different payback categories. Each type of payback on the tick list would be assigned a number of points. The summation of the points for all items in each category provides a quantitative indicator for the amount of payback from research in these categories. If a qualitative description would be supplied along with the items on the tick list, scorers would have some discretion in how many points were awarded for each individual payback.

- **Advantages:** The tick list requires relatively little time to complete; it is easy to fill in and it provides information in easy to use form. When conducted electronically or on-line, the responses can be immediately collected, aggregated and analysed. The catalogued payback ideally relates to discrete and measurable items, such as PhD’s or

publications, that are recognisable to all stakeholders. This will be most straightforward for the first two payback categories: “Knowledge Production” and Research Targeting and Capacity Building”. If combined with a menu list, the responses allow for quantitative comparison within and across the different payback categories. The number of points assigned to the different outputs can be used to reflect the different weights on payback criteria of different stakeholder groups (for example researchers may value paybacks differently to clinicians or patients).

- Disadvantages: Using tick lists may not reflect the full potential of possible research payback. There will always be research outputs that are not included in the tick list. An ‘other’ category may be used to capture these, but if not mentioned explicitly, outputs may not be recognised by researchers. Conversely, in an attempt to capture all possible outcomes, the form may become unwieldy (and yet still may not capture full range of paybacks). Consequently, tick lists will have to compromise on the level of detail. Researchers may not like ‘tick box’ nature as it does not acknowledge the unique character of individual research projects. Furthermore, using points for different outputs (‘menu list’) is subjective in principle and may be too crude. It may be difficult to determine value of each type of output, especially if these values must bear any significance relative to the other outputs.

3.2 Exemplar scales

Scales can be helpful for assessing the payback of arthritis research. In this case we suggest a scale with a number of the scoring levels from hypothetical projects that would score at particular points on the scale. This system provides fixed comparisons for scorers to compare grants against. The assessors’ scores will be based on information in end of grant reports provided by the researcher.

These scales can be used for all payback categories. Some categories, however, will require two or more scales; for example “Impact on health policy” can be split into “level of the policy” (e.g. institutional, national, international) and “degree of impact on health policy (e.g. incremental, considerable, moderate, etc.).

In a previous project we used an 11 point scale with exemplars at the 0, 2, 4, 6, 8, and 10 scores. Our experience suggests there is a tendency of scorers to use only the points on the scale that are defined (the even numbers). An example of this system is illustrated in Box 1; see also Appendix C for examples on other payback categories.

- Advantages: These scales enable scorers to use their personal discretion for identifying and assessing payback that would not necessarily be captured in tick lists. Scales allow for a wide spectrum of payback to be catalogued without the need to specifically identify each output. Exemplars help assessors to improve the consistency of scoring and they provide assessors with a relatively detailed indication of levels. Although our previous experience suggests that there are still numerous issues to resolve around coherence and consistency of scoring, statistical analysis of previous results has

illustrated that the differences between projects with the highest and lowest were significant.⁵

- Disadvantages: assuming that for “Knowledge Production” the scale would simply be a list of examples of numbers of papers/presentations, it is a more subjective version of bibliometric assessment and it can take account of non-ISI publications or presentations. If these scales must cover a wider range of possible outputs, they will have to refer to relatively abstract concepts, such as “considerable impact on health practice”. It may not always be easy for stakeholders to conceptualise specific outputs to these payback scales. Furthermore, the exemplars must be applicable to all areas of **arc** research, or a number of exemplar series must be developed, tailored to different research types or areas. Exceptionally high impact projects may not match with the exemplars, and it will therefore be difficult to recognise them as such.

Box 1. Example of scoring system using exemplars for knowledge production.

Publications
<i>We would like you to consider those publications and presentations that contribute to knowledge dissemination; thus, this implies a judgement of the impact factor and quality of the journals as well as a purely quantitative assessment.</i>
10 – The project produced at least 40 peer-reviewed articles/editorials in respected journals.
8 – The project produced at least 20 peer-reviewed articles/editorials in respected journals.
6 – The project produced at least 8 peer-reviewed articles/editorials in respected journals.
4 – The project produced at least 3 peer-reviewed articles/ editorials in respected journals or journal(s) relevant for the target audience e.g. in a (inter)national professional journal., or at least 4 presentations made at international conferences, or more than 6 presentations made to national meetings.
2 – The project produced internal publications (e.g. report of department); or at least one published abstract or presentation; or at least 2 conference posters or external presentations made, for example to a practitioner, patient or research group (including continuous medical education).
0 – The project produced no publications or internal reports or presentations/posters etc.

3.3 Calibrators

Calibrator scales, though similar to exemplar scales, do not use exemplars from a hypothetical research grant but from actual completed **arc** grants. In each scoring exercise some scores for grants from previous rounds are included. These grants act as ‘calibrators’ allowing the scores in the new round to be normalised to the scores used previously. This system would have a one-dimensional Likert scale for each payback category. Likert scales traditionally use a five-point scale, in which respondents normally indicate their level of

⁵ At 95% confidence interval.

agreement with statements that express agreement or disagreement toward a concept being measured. In this case, the low end of the scale would be defined as ‘No impact’, but no other points defined. The calibrators define the levels of payback measured by the other scales.

- **Advantages:** Rather than comparing to abstract concepts, this tool provides concrete examples to compare against. Furthermore, for these scales there is no risk that the descriptors do not capture the spectrum of possible paybacks, since there are no descriptors. The calibrators define the levels of payback and the scales are thus flexible to the dynamics of research payback. The scales are relatively easy to define as previous impact is the defining factor.
- **Disadvantages:** We expect that the consistency of scoring, using a five-point scale, is lower than for scales with more defined points.⁶ Furthermore, it will be hard to test the validity of this system, except on a larger scale. In order to enable scoring, assessors may also require a lot of information about the calibrator projects. Also, the confidence of the system relies on the assumption that different people will give the same relative project ratings (to the calibrator projects to allow normalisation); this may not be the case. Finally, maintaining the system would require periodic recalibrations of the scales, which adds to the workload of using this tool. One particular danger is that, as the highest scale will be defined by the project with most impact thus far, new projects with a higher impact might require revision of the scale. Such drift in the values of scores over time could make comparison longitudinal analysis difficult.

3.4 Pathways

The tools explained in this and the following sections are conceptually different from those described above. Whereas the previous tools could be used for cataloguing (and quantifying) different paybacks, these tools in this section aim to map the journey to impact from a longitudinal perspective. The effects of research on health care can often be traced in a series of stages. In the pharmaceutical sector for example, most drugs have been discovered either by identifying the active ingredient from traditional remedies or by serendipitous discovery and then follow a well worn pathway through development and testing into use. There are analogous pathways in other areas, for example, in physiotherapy, techniques developed as a consequence of allied health professional research, are validated and disseminated through professional publications and workshops, hopefully leading to widespread adoption.

These tools would ask researchers to map out the pathways along which, and the barriers over which, their research might develop towards patient benefit. They could also be asked to provide an indication of the likely timeframe of potential impacts. The process of drug discovery, for example, involves the identification of candidates, synthesis, characterisation, screening, and assays for therapeutic efficacy. Once a compound has shown its value in these tests, it will begin the process of drug development prior to clinical trials. The routes to impact mapped with this tool could extend beyond the first two payback categories and

⁶ This expectation is based on experience in projects conducted for the Arthritis Research Campaign and for the Netherlands medical research council ZonMW.

into: ‘Informing policy and product development’; ‘Changes in health practice’; ‘Health and health sector benefits’; and ‘Broader economic benefits’.

This tool would involve asking grant holders for an outline flow diagram, bullet points, or narrative, showing the key stages that would be required for the research to affect patients assuming everything went to plan or turned out positive. These pathways can be single and linear, but there could also be multiple parallel pathways with feedback loops. An example of a single linear diagram is given in Figure 2, which illustrates the generalised pathway for medical products in the UK, adapted from the recently published Cooksey report.⁷ Scorers could then be asked to estimate the timescales for each of these stages to provide an estimate of time to impact. This could help to estimate the timescale upon which impacts are expected. Additionally, they could be asked to assess the likelihood of overcoming each of the barriers to reach these potential levels of impact.

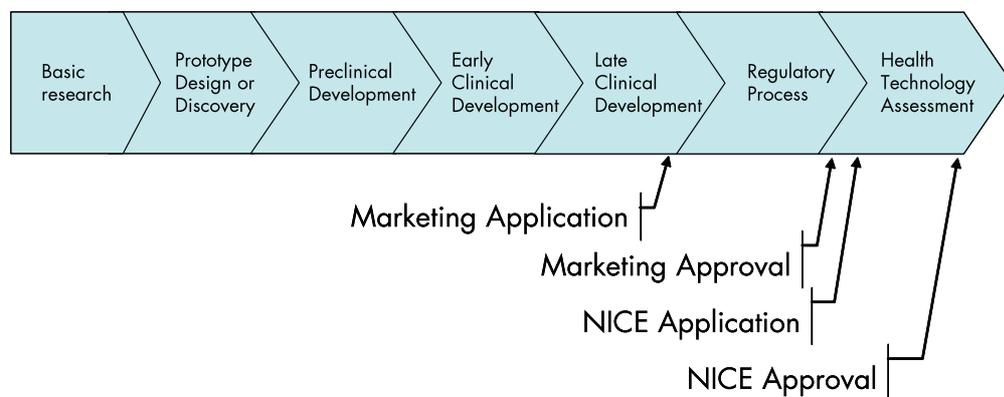


Figure 2. Generalised pathway for medical product. Adapted from Cooksey Report (2006).

- Advantages: Aside from collecting information, this tool makes researchers aware of their role in research facilitating research payback, and encourages them to think about possible ways to achieve this. The assessment can be tailored to specific research characteristics, even for research with (potentially) atypical impact. As these hypothetical pathways assume that all barriers to impact are overcome, this tool may allow potentially important, but unlikely, impacts to be explored. The researchers’ experience and expertise can be used to assess the time to impact and the potential level of impact; however, this tool is relatively labour extensive for assessors.
- Disadvantages: One of the disadvantages of this tool is that it does not result in quantitative indicators for research payback. It will be difficult to aggregate the results of this assessment to the level of **arc**’s research portfolio without scoring or assessment. Also, as research outcomes are, by definition, difficult to predict, it is unknown how difficult it will be for researchers to draw pathways to impact. Furthermore, it is unknown to what extent researchers are aware of the obstacles to research impact.

⁷ Cooksey, D (2006) A review of UK health research funding. Dec. 2006, HM Treasury Independent Review. Available at: http://www.hm-treasury.gov.uk/media/56F/62/pbr06_cooksey_final_report_636.pdf

3.5 Matrices

This tool, similar to the payback pathways tool, would classify the research in three dimensions: probability of impact, size of impact, and time to impact. It could, in fact, be used to summarise the information from a payback pathway. It could be used to identify individual grants that have potential for future development. Also, it would provide **arc** with an up-to-date overview of the predicted impact of their recently finished grants.

As mentioned above, it is difficult to accurately estimate the size of, time to, and likelihood of impact. Consequently we will use crude categories that merely allow distinction of substantial deviations. Researchers would then be asked to provide a statement about the timescale over which their research will eventually feed through to patient benefit: short (less than a year), medium (2 to 5 years), or long (more than 5 years). Similarly, they will be asked to state whether it is likely, possible or unlikely that their research will lead to patient benefit. Finally, the researchers will be asked: if and when their research would have an impact, and would it be small, substantial or revolutionary. Prior to this classification, it would be important to provide an overview or examples of what **arc** considers impact; for example: an important improvement in the quality of life for a population of 1,000 patients might be considered a substantial impact.

To provide a nuanced use of likelihood, timescale and size of impact, the researchers would be requested to fill out an ‘impact profile matrix’. This approach helps researchers to distinguish between, for example: 1) a project with a relatively substantial short-term impact, but it fades out after a few years; and 2) a project with a relatively small short-term impact that sustains and evolves into a substantial impact in the long run. An example of a completed matrix is given in Table 1.

Table 1. Impact profile matrix: example of an estimate of likelihood for different impact sizes and timescales

	Short (= <1 year)	Med (2-5 years)	Long (>5 years)
Small impact	<i>unlikely</i>	<i>possible</i>	<i>likely</i>
Substantial impact	<i>none</i>	<i>unlikely</i>	<i>likely</i>
Revolutionary	<i>none</i>	<i>none</i>	<i>unlikely</i>

These classifications will inevitably be subject to uncertainty as they are based on expert judgement, but by using large buckets to split up the three axes we may be able to provide useful information at an aggregate level. Alternatively, these assessments could be made by external assessors based on the end of grant reports.

- **Advantages:** The main advantage of this mechanism is that the assessment can be conducted at the end of the grant, and submitted with the final report. **arc** will have an up-to-date overview of the predicted impact of their research portfolio. As the researcher is responsible for the statements, the results will be subjective, but this tool does not have the disadvantages of scorer inconsistencies. As the categories are defined relatively broadly, the differences between individual grants are likely to be more significant than in the scale mechanisms discussed above. It also allows **arc** to aggregate the time to, likelihood of and size of impact for their research portfolio, and different subsets thereof. These matrices are particularly useful for later categories in which

payback has not yet be measured; for example “Health and Health Sector Benefits” (and possibly “Informing Policy and Product Development” and Broader Economic Benefits”).

- Disadvantages: As yet, we do not know how accurate predictions of impact are. Estimating and predicting impact is likely to be difficult; the reliability of these statements require critical evaluation. It may also be difficult to design the category ‘buckets’ of the three dimensions so that the descriptions have similar intrinsic attractiveness. Researchers may not be willing to tick a box stating that their research is ‘unlikely’ to have patient benefit.

3.6 Short-listing promising projects

This tool involves a periodic meeting of several experts in research translation to shortlist those projects that have potential for further development. This group of experts would consist of a mix of people with varying fields of expertise and stakeholder perspectives.

Every year, or six months, this group would convene to review the **arc** end of grant reports and assess and discuss the potential of these studies. This development mainly focuses on the payback categories: Informing policy and product development; Changes in health practice; and Health and health sector benefits. Promising **arc** grants will be short-listed using a set of criteria, for example:

- Importance of the patient group, or the size of the patient population in Britain. For example, back pain will score high on this criterion, because there is a potentially large affected patient population.
- Attractiveness to practitioners. This criterion involves the simplicity of intervention and feasibility of implementation. If the intervention involves an obscure piece of kit, it may get low priority. If practitioners could immediately apply the intervention without any additional training or instruments, it will get a higher priority.
- Attractiveness to patients. The simplicity of intervention and feasibility of implementation is also important from a patient’s perspective. If the intervention has a positive or a small negative impact on the daily routine of a patient, it will get a relatively high priority. A daily exhaustive exercise for 2 hours using an expensive instrument, for example, will have a relatively low priority.

The group should have a generic responsibility for all grants, through representatives of basic and clinical research, allied health professionals, patient groups, and possibly the pharmaceutical sector. Alternatively, the group could have subcommittees responsible for grants in specific fields. There will be strict guidelines for avoiding conflicts of interest, or at least, making them transparent.

- Advantages: The short-listed projects are the result of a consensus-based process and reflect the perspectives of the stakeholders involved. Bringing together all relevant expertise, **arc** and interested parties such as the pharmaceutical industry are periodically presented with a list of promising projects. The overview will be compiled shortly after the grants have been finished. There will be no additional commitment required from the researchers; they are only asked to submit the end of grant report.

- **Disadvantages:** The group will have to speculate about potential impact; there is no estimation of likelihood or timescale. As this tool focuses on identifying opportunities, it can not be used by **arc** for management decisions. It will not be possible to aggregate and compare scores across fields of research or types of grants. The shortlist is a subjective decision of a group of experts, who may not necessarily be in the best position to judge the potential of the research; the researchers themselves are not involved. It does not seem to be very applicable to basic research, as potential treatments or products are yet difficult to predict.

3.7 Searchable repository

This tool attempts to bridge the gaps between academic researchers producing knowledge about new therapies or potential new targets and industry implementing these discoveries. Hitherto, the main commodity for knowledge production is still academic public. The trouble with publications, though, is that they are produced only if the project is successful and then possibly several years after completing the work. Therefore, industry identifies academics who are working in areas of their interest by networking and from conferences and publications.

To facilitate this process, **arc** could develop a searchable repository of **arc**'s research available on its website for pharmaceutical companies to identify opportunities for drug development, or for research collaboration. This repository would be populated with periodic updates of research abstracts from end-of-year reports or even grant applications. Hence, the research available through this repository would be different from the abstracts available through publications' databases, such as PubMed. The grants could be categorised by key words (e.g. animal model, cartilage breakdown), by methodology (e.g. animal models or in-vitro analysis), and/or by disease area (e.g. osteo-arthritis). This way, the pharmaceutical companies would not have to rely on a short list produced by **arc** experts, they could use their own criteria to select promising research results.

- **Advantages:** Tailored to the needs of the pharmaceutical industry; companies can use their own set of criteria when short-listing promising projects. The project can be available in the database immediately after the end of year report has been received or the grant has been awarded. There will be little additional commitment required from the researchers; they are only asked to submit the end of grant report (except suggesting key words). Such a database could also be used by academic researchers to identify collaborators. As there is no assessment involved leading to arbitrary decisions, there are fewer possible grounds for disagreement between stakeholders.
- **Disadvantages:** The repository has only limited purpose and will not fulfil all of **arc**'s needs. Impact cannot be assessed or predicted with this tool and it does not provide **arc** information on likelihood, size or timescale of payback. It cannot be used to support management or strategic decisions. And the repository is only potentially useful for product development; for example, policy making will not be informed by this tool.

In general, currently **arc** requires end of year reports for all of its funding. It also requires end of grant reports for project, programme and fellowship funding. These reports are relatively unstructured and little use is made of the information in them. We would hope to minimise additional workload on researchers by focusing on collecting only information that will be used and possibly reducing reporting frequency. **arc** also asks for some information about the outcomes of previous grants when applicants apply for new grants, but again this is not aggregated.

The end of grant report, however, is not an ideal moment for the collection of information on research outcomes, as there is a substantial time lag associated with finishing research activities and payback of research outcome. This holds especially for outcomes related to the last three payback categories (Informing Policy and Product Development; Health and Health Sector Benefits; and Broader Economic Benefit). Therefore, it is likely that certain research payback occurs after the moment of cataloguing. Therefore, in addition to actual impact, the system could also account for predicted impact, the outputs that can be expected, based on extrapolation of the outputs to date.

There are various options for when to collect information, or a hybrid of these techniques could be used:

At start of grant

Initial information about likely paybacks could be collected at start of grant or even in the application phase, this could then be updated at end of grant.

- **Advantages:** there is an incentive for researchers to supply information at the moment of application that is adequate and timely. Researchers are likely to more compliant when their research funding is at stake.
- **Disadvantages:** actual outcomes cannot be measured, and it may be difficult to accurately predict outcomes before research has been conducted, And if information is collected in the application phase, there is an unnecessary administrative burden on researchers whose submission will be rejected.

At end of grant

At the end of the grant, information can be collected about actual paybacks that have resulted from the research.

- **Advantages:** slightly better idea of paybacks and results of research is known.

- Disadvantages: when measuring actual output, little payback is likely to have occurred by end of grant; it may still be necessary to estimate expected payback. Previous experience has strengthened the evidence about the difficulty of assessing and scoring impact too soon after the completion of the projects. Furthermore, it may be difficult to get researchers to engage in the process; there is little incentive for them to actively commit to collecting information on research outcomes,

It is also possible to collect information at end of the grant about expected payback, using predictions of impacts in the following 2 or 5 years time.

- Advantages: payback is more likely to occur in the 2-5 years following the research. Therefore, such information will better represent the payback arising from research grants, especially at the level of **arc**'s entire research portfolio.
- Disadvantages: the confidence of expert judgement about expected impact is debatable.⁸ Verification and subsequent calibration of the estimates 2-5 years after the estimates will be required, which would involve double measurement.

After end of grant

From a research evaluation perspective, collecting information would be most valuable at the moment when actual payback can be measured. That would be between 2 and 10 years after the research grant has ended. While 10 years would not be feasible, **arc** could follow up their grant holders 2 to 5 years after completion and measure actual payback.

- Advantage: it is more likely that actual payback has occurred, rather than estimating expected payback.
- Disadvantages: there is little incentive for researchers to engage in collecting this information. Incentives would be required but not necessarily sufficient. For example contractual obligations or financial incentives.

On application for next grant

An opportunistic approach would be to encourage researchers to supply information on payback of previously held grants when they apply for the next one.

- Advantages: longer timescales allow more distant paybacks to be catalogued. There will be a possibility to monitor impact over time if researchers are required to report on every previously held grant every time they apply for a grant at **arc**. Furthermore, there is an incentive for researchers to engage, as their research funding will be at stake.
- Disadvantages: not all previous grant holders will apply for new grants at **arc** (although most do) and therefore the coverage of payback cataloguing will not be complete⁹. Timescale of assessment are not predictable, although given the frequency of re-application collecting information from only grants within certain time windows is feasible.

⁸ In several previous articles Buxton and Hanney have been critical of previous work by, for example NIH and Detsky, that takes research findings and makes assumptions about the expected levels of payback, unless it is done as a benchmark for subsequent studies after 3-5 years. This would involve a double assessment.

⁹ Analysing historical grant data from 1987-2006 in over 80% of cases one applicant for a grant will re-apply for funding in the window 1-3 and 5-7 years after the grant completes.

Sampling

Using predictions of expected impact at the moment of the end of the grant, **arc** could return to (a sample of) older grants after 2 or 5 years to update payback catalogue.

- Advantages: this provides information after paybacks have had a chance to mature, could act to calibrate predictions of payback
- Disadvantages: there is little incentive for researchers to engage, and it would require additional work to identify and follow up researchers

Combination of approaches

The above methods could be combined – for example researchers could be asked to evaluate paybacks at the end of grants and then asked to update these assessments when they applied for additional funding. If they did not apply for additional funding within 5 years of their grant finishing **arc** could approach them to update their assessment at the 5 year point.

- Advantages: This approach allows us to cherry-pick the best elements from the different options and combine their advantages, while mitigating disadvantages
- Disadvantages: a combination of approaches adds considerably to the complexity of the system.

APPENDICES

Appendix A: Key Informant Interviews

Perspective	Name	Affiliation
Basic research	Prof Gillian Murphy	University of Cambridge
Clinical research	Prof David Reid Prof Peter Croft	University of Aberdeen Keele University, Institute of Primary Care & Health Sciences
Industry	Dr Andrew Docherty, vice President of Research and Alliances Melissa Rubock, Academic Liaison Dr Adrian Moore, Senior Pharmacologist	UCB Group UCB Group UCB Group
Policy	Dr Maxime Dugados	Hopital Cochin, Service de Rhumatologie B
Research funder	Fergus Logan Mike Patnick Prof Alan Silman	Arthritis Research Campaign Arthritis Research Campaign Arthritis Research Campaign

Appendix B: Draft tick lists

In this appendix we present a draft Tick List for research outputs and outcomes of **arc** funded research. This list was compiled from our previous work examining the long term outcomes of **arc** funded research.

KP

Type of knowledge

Improved understanding of the normal state

Improved understanding of the diseased state

Improved understanding of the identification of the diseased state

Improved understanding of the mechanism of disease

Improved understanding of the prevalence of disease

Improved understanding of the risk factors

Improved understanding of possible interventions into the disease state

Improved understanding of current interventions into the disease state

Proxies of size of contribution

Papers in peer reviewed journals

Impact factor >5

Impact factor <5

Presentations at peer reviewed conferences

Posters at peer reviewed conferences

Book chapters

Books

RTCB

Research influence

Influenced own research

Follow on grants (>£200K)

From **arc**

From other funders

Influenced research of others

Researchers in academia

Researchers in industry, or industrial organisations

Researchers in policy

Training

Research has contributed to training or advancement of researchers

Qualifications

Ugrad research projects

Masters

MDs

PhDs

Fellowships

Chairs

Career

Established new research group, ie stepping stone to independence

Sustained research group

Increased size of research group

Improved capacity to use research

By practitioners

- By policymakers
- Tools for research
 - New markers or screening factors for disease identified in
 - Cell culture
 - Animal model
 - vivo
 - Observational
 - New models of disease discovered in
 - cell culture
 - animal model
 - New assay for screening factor
 - cell culture
 - animal model
 - in vivo
 - observational
 - New target identified in
 - cell culture
 - animal model
 - vivo
 - New target validated in
 - cell culture
 - animal model
 - vivo
- IPPD
 - Proxies for extent
 - Citations in clinical guidelines
 - Local/Regional
 - National
 - European/International
 - Citations in systematic reviews
 - Interactions outside research
 - Briefings to policymakers
 - Local/Regional
 - National
 - European/International
 - Briefings to practitioners
 - Local/Regional
 - National
 - European/International
- Newspaper articles
 - International
 - National
 - Regional/local
- Industrial links
 - Advice provided to industry
 - MTAs with industry
 - Consultancies with industry
 - Co-authorship with industrial researchers
- Product Development
 - Patents
 - Licenses
 - Therapeutic Products
 - in animal tests
 - in phase 1 clinical trial
 - in phase 2 clinical trial
 - in phase 3 clinical trial
 - on the market
 - Non pharmaceutical interventions
 - eg exercise therapy
 - proof of concept
 - large scale trial
 - wide spread use
 - Diagnostic Tests
 - proof of concept
 - prototype
 - on the market

Appendix C: Example exemplar scales

Based on categories in the payback framework and earlier work by HERG and RAND, we developed a system using exemplar scales to score research outcomes. As we would also like you to consider application of the findings by health practitioners on (inter)national or local level, we included an additional category related to the changed behaviour of practitioners and managers.

The levels on each scale range from 0 (no impact) to 10. To assist scoring the exemplar descriptors for the levels for each scale are meant to indicate the type of impact that would be expected at each level but clearly there has to be room for interpreting the diverse situations that will arise. This is further facilitated by providing exemplars for scores 2, 4, 6 8 and 10 only. There would of course also be an “I don’t know” option for each scale.

For this draft exercise some indications about appropriate ceiling level scores for each scale were taken from the top examples from the 16 case studies conducted in previous work for **arc**, which included examples of the most productive **arc**-funded research, especially in terms of publications.

1. Knowledge Production

1a. Publications

We would like you to consider those publications and presentations arising from the research in question that contribute to knowledge dissemination; thus, this implies some judgement of the quality of the journals as well as a purely quantitative assessment.

- 10 – The project produced at least 40 peer-reviewed articles/ editorials in respected journals.
- 8 – The project produced at least 20 peer-reviewed articles/ editorials in respected journals.
- 6 – The project produced at least 8 peer-reviewed articles/ editorials in respected journals.
- 4 – The project produced at least 3 peer-reviewed articles/ editorials in respected journals or journal(s) relevant for the target audience e.g. in a (inter) national professional journal. or at least 4 presentations made at international conferences, or more than 6 presentations made to national meetings.
- 2 – The project produced internal publications (e.g. report of department); or at least one published abstract or presentation; or at least 2 conference posters or external

presentations made, for example to a practitioner, patient or research group (including continuous medical education).

- 0 – The project produced no publications or internal reports or presentations/ posters etc.

2. Research Targeting and Capacity Building

2a. Capacity building

We would like you to consider the development of research skills, personnel and research capacity, especially as evidenced in research careers or degrees.

- 10 – The project made a considerable contribution to at least 4 research degrees, such as PhDs, and to the careers of at least 2 of the other researchers.
- 8 – The project either made a considerable contribution to the careers of at least 2 of the researchers, or to 2 research degrees, or a moderate contribution to at least 4.
- 6 – The project made a moderate contribution to at least 3 research degrees or careers.
- 4 – The project made some contribution to at least 2 research degrees or careers.
- 2 – The project made some contribution to at least 1 research degree or career.
- 0 – The project made no contribution to research degrees or research capacity building in any other way.

2b. Research targeting

We would like you to consider the contributions made by the research to follow-on research: for example, informing replication studies, pilot studies or implementation projects; providing methodological frameworks or data for further studies; highlighting unresolved issues etc.

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

3. Informing Policy making and Product Development

We would like you to consider policy making at (inter)national or local level of the health system, e.g.: reimbursement decisions by NICE; citing of the findings in an advice from the NHS in a NSF or decision by the National Screening Committee; citing of the

findings in a clinical guideline from a national or local professional group; inclusion of the findings in a contract or in a document from an audit, an inspectorial or an evaluative body; the establishment of a working group to examine the implications or implementation of the findings etc.

3a. Nature/level of the policy

- 10 – The project made an impact on a substantial policy of an international body or several substantial policies of at least the national government.
- 8 – The project made an impact on at least one policy from a national body such as the NICE or NSF.
- 6 – The project made an impact on at least one policy from a national professional body.
- 4 – The project made an impact on the policymaking at a sub-national level.
- 2 – The project made an impact on the policymaking of a group of professionals at at least one local unit of the health service.
- 0 – The project made no impact on policies.

3b. Degree of impact on health policy

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

3c. Informing product development

We would like you to consider products as widely defined to include drugs, devices etc.

- 10 – At least one major product (drug, device etc) was considerably based on the research findings.
- 8 – At least one major product was moderately based on the research findings, or several products were considerably based on the findings
- 6 – At least one major product was based to some extent on the findings, or several products were moderately based on the findings
- 4 – At least one product was moderately based on the findings
- 2 – The findings played some role in the development of at least one product
- 0 – No impact on product development.

4. Health and Health Sector Benefits

We would like you to consider application of the findings by health practitioners on (inter)national or local level, such as: the application of policies; changed behaviour of practitioners & managers; the involvement of health service users.

4a. Level of impact in health practice

- 10 – The project made an impact on behaviour in more than one country.
- 8 – The project made an impact on behaviour at national level.
- 6 – The project made an impact on the behaviour at a sub-national level.
- 4 – The project made an impact on the behaviour of at least one team of practitioners or managers, or at least one group of patients/ members of the wider public.
- 2 – The project made an impact on behaviour of at least one or more practitioner, manager, patient or member of the public.
- 0 – The project made no impact on behaviour.

4b. Degree of impact in health practice

- 10 – At least one of the changed behaviour(s) was almost solely based on the project's evidence in a direct instrumental way.
- 8 – The project made a considerable impact on the behaviour(s).
- 6 – The project made a moderate impact on the behaviour(s).
- 4 – The project made some identifiable impact on the behaviour(s).
- 2 – A claim for impact made but limited details given.
- 0 – No impact on behaviour.

We would like you to consider the broader (societal) impact on health (ie the health gain which is a combination of the amount of impact on any one individual and the potential number of patients involved) and/or on the health sector (eg improved efficiency/cost reduction in the delivery of existing services; qualitative improvements in service delivery; greater equity through improved allocation of resources at an area level, improved geographical accessibility, improved equity of access on a gender or ethnic basis etc).

4c. Patient benefits and/or system benefits

- 10 – The project made a very considerable impact on health in terms of many patients making a considerable health gain, **and possibly** on the health sector through the other factors listed;
- 8 – The project made a considerable impact on health in terms of either patient numbers or the health gain of each individual, **and possibly** on the health sector through the other factors listed;
- 6 – The project made a moderate impact on health in terms of either patient numbers or the health gain of each individual, **or** on the health sector through the other factors listed;

- 4 – The project made some impact on health in terms of either patient numbers or the health gain of each individual;
- 2 – A claim for impact made on the health sector in relation to the other factors listed but limited details given;
- 0 – The project made no impact on health/health sector.

4d. Degree of impact on health benefits

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

5. Broader Economic Benefits

5a. Broader economic benefits from commercial exploitation

We would like you to consider the broader economic benefits from commercial exploitation in terms of factors such as increased employment, exports, import substitution.

- 10 – The commercial exploitation based almost entirely on the findings has led to a considerable level of employment, exports and import substitutions.
- 8 – The commercial exploitation based to a considerable extent on the findings has led to a considerable level of employment, or exports or import substitutions.
- 6 – The commercial exploitation based at least moderately on the findings has led to a moderate level of employment, or exports or import substitutions.
- 4 – The commercial exploitation based at least partly on the findings has led to a moderate level of employment, or exports or import substitutions.
- 2 – The commercial exploitation based at least partly on the findings has led to some employment, or exports or import substitutions.
- 0 – No evidence of broader economic benefits from commercial exploitation.

5b. Broader economic benefits from a healthy workforce

We would like you to consider the broader economic benefits in human capital terms as reflecting the value of the lost production avoided as a result of reduced mortality and morbidity.

- 10 – A considerable increase in the value of production no longer lost through premature mortality and morbidity as a result of findings that almost entirely led to the health gain.

- 8 – A considerable increase in the value of production no longer lost through premature mortality and morbidity as a result of findings that led to the health gain in a considerable way.
- 6 – A moderate increase in the value of production no longer lost through premature mortality and morbidity as a result of findings that were moderately responsible for the health gain.
- 4 – Some increase in the value of production no longer lost through premature mortality and morbidity as a result of findings that were moderately responsible for the health gain.
- 2 – Some increase in the value of production no longer lost through premature mortality and morbidity as a result of findings that were in any way responsible for the health gain.
- 0 – No evidence of broader economic benefits as a result of a healthy workforce