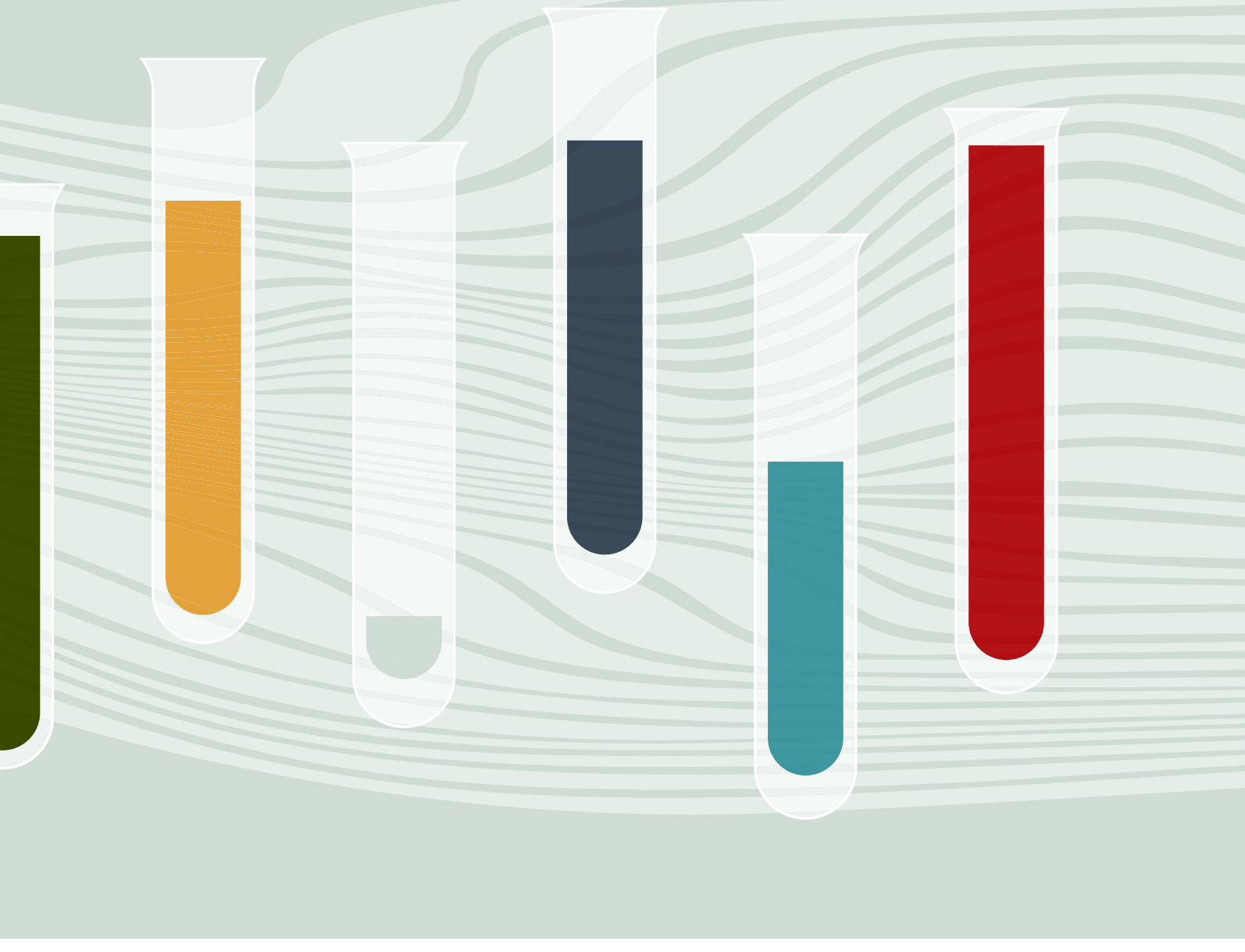
New Frontiers in Evaluation of Impacts of Medical Research

International Workshop 2009



Vetenskapsrådet

VETENSKAPSRÅDET
SWEDISH RESEARCH COUNCIL

Box 1035
101 38 Stockholm
SWEDEN
www.vr.se

Print: CM Gruppen, Bromma, Sweden, 2009

Writer: Lynette Gilbert
Graphic Design: Erik Hagbard Couchér

New frontiers in evaluation of impacts of medical research

Governments around the world have been facing increasing demands for greater accountability and efficiency of their public investment into research. Until recently health and medical research has been insulated from these pressures, but are now very much part of international benchmarking practices. There is also increased demand for research councils working in these fields to become more accountable.

In order to retain credibility amongst the public and politicians, there needs to be an improvement in how the need for medical research funding is promoted, as well as how this research impacts society.

In November 2007, the Scientific Council of Medicine within the Swedish Research Council brought together a small international group of evaluation practitioners in Sigtuna, Sweden. Their aims were two-fold: to identify better ways to measure the impact of medical research investment and to help research funding bodies make a stronger case for funds from government and elsewhere.

The workshop reached two conclusions: Firstly, evaluators need more accurate ways to estimate economic returns, and secondly, greater international collaboration is required to advance knowledge on crucial issues.

These issues include understanding how innovation take place, how to best analyse the social and cultural impacts of research and how research outcomes can be attributed to individual funders.

A working group was set up, led by Dr David Cox of the UK Department of Health, tasked with creating a roadmap that defines key questions to explore and possible approaches to take. Other members of the group were Gerrit van Ark at ZonMw, Netherlands, Peggy Borbey of the Canadian Institutes of Health Research, Martin Buxton at Brunel University, England, Per Carlsson at Linköping University, Sweden, Susan Cozzens at Georgia Tech, USA, Jonathan Grant of RAND Europe, England, and Toni Scarpa of the National Institutes of Health, USA.

In May 2009, a second Sigtuna workshop met to review progress over the intervening two years and to debate new challenges. The participants included researchers, representatives from funding organisations, policymakers and evaluators.

The Scientific Council for Medicine cannot stress enough the importance of addressing methods or parameters on an international level. The Sigtuna 2 workshop brought all these preceding discussions together.

It then moved the debate forward onto how research funding agencies and stakeholders can better understand agency performance. Such information is key to better strategy development and implementation. The workshop focused on what issues can be addressed now with current evaluation practices and what conceptual and methodological questions should be top of the agenda for future work.

The discussions at the workshop, and the work of the core working group, show that it is not an easy task to measure the impact and outcomes of medical research. However the fact that it is difficult does not make it less essential. It is of great importance both for the financing organisations and politicians, which also makes it important for academia. I hope that the work of trying to agree on different approaches in measuring returns will continue.

The key themes and suggestions from the Sigtuna 2 workshop are summarised in this booklet by writer Lynette Gilbert.

I would especially like to thank the core working group for all their important work, input and exchange of ideas. We will publish a report of their work, as a complement to this booklet.

Håkan Billig
Secretary General, Scientific Council for Medicine
Swedish Research Council

MORE INFORMATION

On the Swedish Research Council website, www.vr.se, you can find:

- Presentations from the workshop in 2009.
- Booklet of the Sigtuna workshop 2009 in English.
- Booklet of the Sigtuna workshop 2007 in Swedish.

Contents

From Advocacy to Action: Evolving evaluation objectives

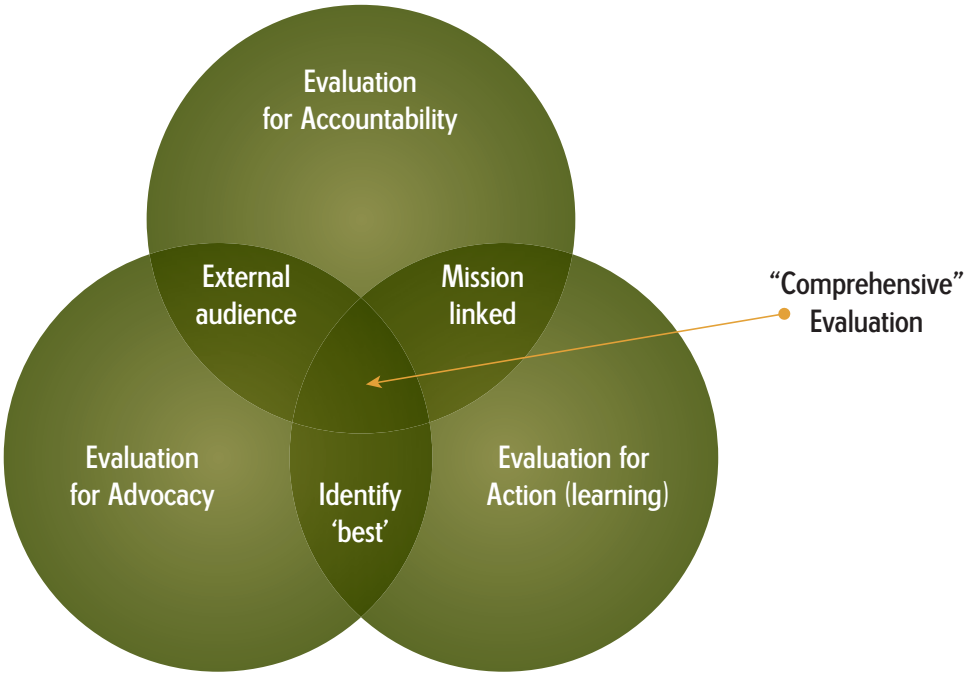
"Our job is to provide a practical evidence base for health science policy, not to select success stories."

[Jonathan Grant, RAND Europe]

The positive impact of medical research has traditionally been assumed. New breakthroughs to cure disease or improve healthcare are welcomed by the public and used by scientists to justify further health research expenditure. However, as the sums invested rise, funding bodies are increasingly seeking evidence that their research spend provides good value for money and acceptable returns. With governments facing tough choices on public spending, medical research cannot rely on anecdote but must demonstrate real economic and social impact if it is to compete for funds against other priorities.

In the face of these demands, doing nothing is not an option. Dr Jonathan Grant of RAND Europe outlined the three pressing reasons for evaluating medical research.

THREE REASONS FOR EVALUATING RESEARCH



Source: Making an impact, Canadian Academy of Health Science

Advocacy: raising awareness , building goodwill

Health is a topic that interests almost everyone, but health research is often complex and protracted. The direct benefits of a particular project or research portfolio may not emerge for some years – the average time lag from grant to health impact is estimated to be 17 years, and is often longer. The challenge for funders is to build a better understanding of biomedical science, research timeframes and the scientific process among policymakers, charitable donors and the general public. Evaluation has an important role to play in providing engaging, objective evidence of impact in areas that genuinely matter to these stakeholders – not only the possible healthcare outcomes, but the potential to create new skills, jobs and investment, or to empower communities in other ways. Research impacts should not simpy be a dialogue between funders and researchers.

Accountability: demonstrating good governance

Funding agencies face pressure from their stakeholders to show that they are making the best possible use of the funds entrusted to them. By adopting systematic mechanisms to monitor the impact of research projects or programmes, funders can start to answer questions such as: Did the research result in 'new knowledge' or change our understanding of health or disease? Have the results of the research been incorporated into clinical guidelines or other guidance? Has the research led to changes in clinical practice or health policy? Did it result in new products, patents or other commercial outcomes? Has the research generated (whether directly or indirectly) other economic, social or cultural benefits?

Action: doing things better

As the evaluation debate continues to move forward, there is growing recognition that the ultimate objective must be to improve the way we fund and conduct research, not simply to monitor and measure. By aggregating evaluation results across their research portfolios, funders can inform their future funding and governance decisions by gaining insights on the effectiveness of different funding mechanisms, or by identifying success factors in the translation of research into practice. Comparing different research approaches and sharing the results with governments, institutions and local decisionmakers allows new interventions – and the context for those interventions – to be adjusted based on evidence of what works. This formative, learning agenda includes questions such as: What are the characteristics of research discoveries that have led to breakthroughs in the diagnosis, prognosis and/or treatment of disease? How are grants actually used – salary, running costs, equipment etc? How do different institutional arrangements and policies affect life science research, health care innovation and health care outcomes? How do different training, organisation and financing models of healthcare influence innovation and utilization patterns? How do research results influence general awareness of lifestyle factors for health, and how does this influence disease prevention behaviours? Having credible, evidence-based answers to these questions would provide valuable strategic insights to inform policy decisions. They would also assist researchers wanting to increase the likelihood that their research project will ultimately have real, practical impact.

Matched to need

Whether the evaluation assesses project-level impacts or international comparisons, objectives drive the research methodology and choice of performance indicators. The UK government's FABRIC checklist is a useful hygiene test at this next, more detailed level – the evaluation approach must be:

- Focused on the organisation's aims and objectives
- Appropriate for the stakeholders who are likely to use the information
- Balanced to cover all significant areas of work performed by an organisation
- Robust enough to cope with organisational changes
- Integrated into management processes, and
- Cost-effective: balancing the benefits of the information against the costs of collection

Objectives should also be capable of being flexible. A change in context or unexpected results may mean something different needs to be measured. A 'failed' project may – viewed from a learning perspective – provide valuable insights on what to avoid, or new methodologies for other purposes. Success and failure may need to be redefined, as the evaluation agenda moves from being summative (description) to formative (performance). •

Simplicity versus Sophistication: Evaluation frameworks compared

"How we choose to measure impact will determine the kind of impact we find."

[Claire Donovan, Australian National University]

The underlying processes of knowledge creation and translation are inherently complex and hard to measure. Multiple research efforts may contribute to one scientific advance. Time lags to impact are hard to predict, and final outcomes may depend on other factors such as local context. To add to the complexity, funders must also make choices on what they wish to evaluate before they can select an appropriate framework.

Selecting a framework: a hierarchy of choices

Philipp-Bastian Brutscher of RAND Europe and Dr Gerrit van Ark of Zon Mw [1] presented separate analyses highlighting the key choices required to fit frameworks to need:

- **Objectives**. Three key objectives are accountability (efficient use of funds), action (to help steer research or guide future allocation decisions) and advocacy (signalling ability). The choice of objective influences evaluation questions, which should reflect the funder's mission.
- **Level of aggregation** may be low (individual researcher, institution or project), intermediate (faculty or programme) or high (research discipline or funder portfolio). Higher levels of aggregation require longer time horizons.
- **Target groups** for the evaluation results, for example politicians, scientists, public or patients.

Governed by these choices, evaluators can then consider detailed methodological issues:

- **Indicators and measures** encompass outputs (e.g publications), outcomes (e.g. clinical practice guidelines) and impacts (long-term changes), which may be scientific, social/health or economic.
- **Timing** may be 'longitudinal', tracking inputs, outputs, outcomes and impacts of one project or programme over time. A 'cross-sectional' perspective looks at multiple projects and outputs within a given timeframe.
- **Methods** fall into 3 broad categories: statistical analysis, modelling, and qualitative and/or semi-quantitative methods.

Different frameworks combine elements in different ways. The Payback framework has an accountability objective, a range of indicators, low to intermediate aggregation, a short (longitudinal) timeframe and employs a few qualitative and semi-quantitative methods. By comparison, the Swedish Vinnova framework has allocation and advocacy objectives, measures long term impacts at high levels of aggregation, has a long (also longitudinal) timeframe and uses many different methods.

Both presenters emphasised that the key choice is that of evaluation objective. Philipp Brutscher described a hierarchy of choices: objective influences choice of indicator(s), which in turn influences aggregation level and timing. Methodology depends on desired level of aggregation: low level aggregation is possible with few methods, but high aggregation typically requires multiple methods. Accountability and/or advocacy objectives are best met by frameworks that combine upstream measures (e.g. outputs), low aggregation and a short timeframe. Action objectives are better served by combining downstream measures (e.g. outcomes and impacts), high aggregation and a long timeframe. Dr van Ark commented that cross-sectional frameworks such as Sci-Quest provide a particular focus on communication with societal user groups.

Capturing social and economic benefits

There is growing emphasis on capturing the broader social and economic impacts of research – six of the eight frameworks analysed by RAND include social and economic outcomes. This does not mean the science is ignored; for example, the LUMC [2] framework focuses on societal impact, while scientific quality is assessed separately by the University's Centre for Science and Technology Assessment.

Several funders have enhanced existing frameworks to create a tighter link to objectives. Two national funders, the Canadian Institutes for Health Research (CIHR) and ZonMw, have created supplementary frameworks that focus on innovation and knowledge translation within their research programmes. Dr Ian Graham described CIHR's 'knowledge to action cycle' which views knowledge translation as a process, not an event or outcome [see graphic]. In addition to end-of-grant knowledge dissemination, CIHR encourages integrated knowledge translation activities and programmes. These engage potential knowledge users to help shape the research, interpret findings and move research results into real-world applications. ZonMw also puts a strong emphasis on implementation and encourages practitioner and patient involvement. Dr Janna de Boer presented the 'total innovation cycle' which underpins ZonMw funding and evaluation. A key indica-

[1] The Netherlands Organisation for Health R&D

[2] Leiden University Medical Centre

tor of performance is whether a project has moved to the next phase of the cycle. Programmes which cover several phases qualify for funding from both health and scientific research sponsors.

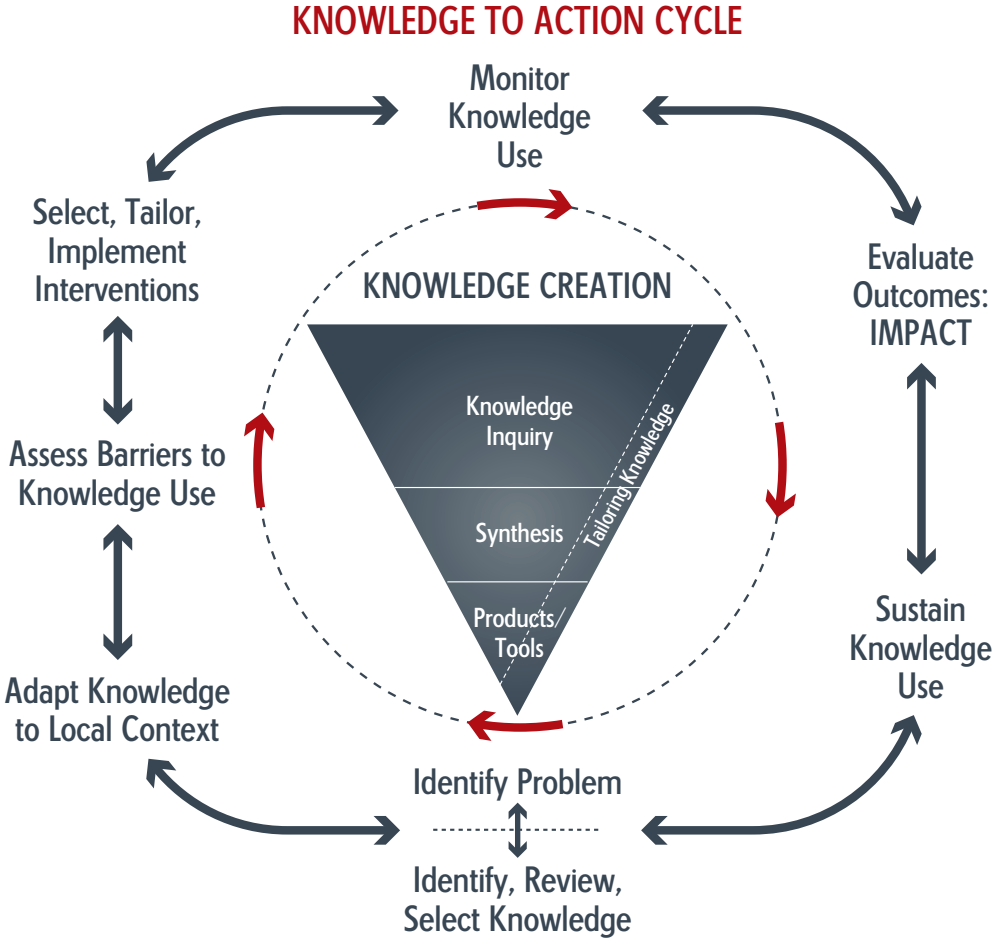
At international level, the World Health Organisation (WHO) plays a facilitative role, working with member states and partners to improve health outcomes across six diverse regions. Robert Terry explained how WHO's 'research for health' strategy will help member states set research priorities and improve their national research capacity, standards and knowledge translation. A WHO priority-setting framework identifies key steps to follow from scoping a health issue, understanding causes, developing solutions, implementing and evaluating. A separate evaluation framework assesses impacts against goals and inputs – evaluation can impact the process at any point.

Simplicity or sophistication: an unfulfilled experiment

Evaluators and funders – quite reasonably – seek transparent and cost-efficient evaluation approaches and simple metrics. However, Dr Claire Donovan warned that oversimplification can yield disappointing results that do not credibly link research funding to research outcomes. Evaluators should be willing to set ambitious goals and embrace greater methodological complexity. Dr Donovan described an Australian initiative to create a world-leading Research Quality Framework (RQF) based on a 'quadruple bottom line' of social, economic, environmental and cultural impacts which would allow application across multiple research fields. 'Transformational' impacts on industry, business and community would be assessed through mixed quantitative and qualitative methodologies, including case studies, context statements, peer review and end user evaluation. These ideas were never tested as, following a change of government in 2007, a less costly – but less insightful – framework was adopted, based on simple impact metrics such as patents and commercial income. The proposed RQF impact assessment methodology has, however, been adopted for the UK's Research Excellence Framework (REF).

One 'super-framework'?

Views differ on whether one 'super-framework' could meet all possible needs, or whether the evaluation community should consolidate on a handful of existing frameworks. A new framework has recently been proposed for use by any Canadian health research funder – see Making an Impact, p 18. ●



Source: Canadian Institutes for Health Research

Getting it done: Evaluation systems and processes

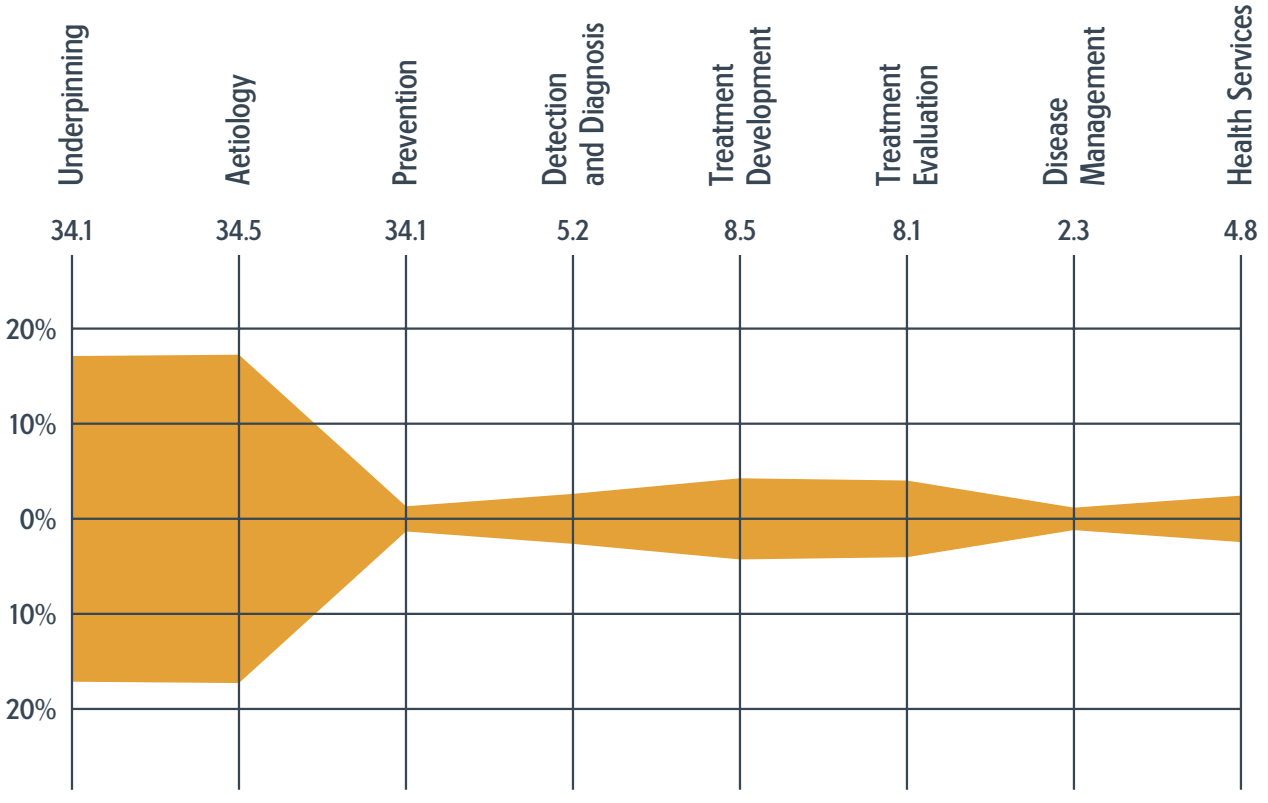
"The perfect is the enemy of the good. We just set out a way to capture something useful."
[Liam O'Toole, Clinical Research Collaboration, UK]

Evaluation is a practical activity. Data must be collected and analysed systematically, and the results shared with researchers, funders and others. Several Sigtuna speakers described the ongoing evaluation systems they use.

Establishing a common baseline: The Health Research Classification System

A pre-requisite to comparative evaluation of research is a consistent approach to classifying projects. Funders have typically developed their own customised classification systems, and this diversity is now a practical obstacle to strategic comparison across research portfolios. The Health Research Classification System (HRCS) is a potential solution to this problem, developed by the UK Clinical Research Collaboration which seeks to improve coordination among major funders. HRCS classifies clinical research along two dimensions: (i) health categories (21 health and disease categories, based on WHO classification codes) and (ii) type of research activity (48 codes grouped into 8 categories, based on the cancer Common Scientific Outline). The results show the 'centre of gravity' of research spend [see graphic]. HRCS is now used by 22 UK organisations and has informed national policy discussion resulting in £50m of joint funding initiatives. It is also being used in some other countries. Dr Liam O'Toole described the process of implementing HRCS across multiple organisations: providing training, showing how HRCS complements existing coding systems and frameworks such as Frascati, establishing a QA approach and holding an international workshop to share lessons. As use grows, governance mechanisms will be needed to ensure HRCS can evolve to meet new user needs without losing integrity.

PROPORTION OF COMBINED TOTAL SPEND BY RESEARCH ACTIVITY – KITE DIAGRAM



Source: UK Health Research Analysis 2006

Rolling evaluation programmes at LUMC and CIHR

Organisations that evaluate multiple programmes need disciplined systems. Professor Eduard Klasen of Leiden University Medical School (LUMC) and Peggy Borbey of the Canadian Institutes of Health Research (CIHR) outlined the approaches used by their institutions. Canadian government policy requires all CIHR-funded programmes to be evaluated for relevance to Canadian priorities and performance. With around 150 active programmes clustered into 22 categories, CIHR evaluates each category every 5 years, and also reports to Parliament annually on overall performance. The results provide evidence for allocation decisions and new learning insights. A strategic expenditure review of CIHR's portfolio also takes place every 4 years, with the lowest performing 5% of expenditure identified for reallocation. A small Impact Assessment unit has been working on theme-based impact assessments of selected priority research areas (e.g. cardiovascular disease) and topics (e.g. commercialisation). At LUMC, some 80 research programmes are evaluated annually by the LUMC Science committee. This looks retrospectively at numbers of theses completed, publications, commercial income earned etc. An additional programme of 3-yearly self-evaluation combines retrospective and forward perspectives, coupled to inputs such as funding and FTEs. Programmes also have an external evaluation visit every 6 years from an independent committee of foreign peers. LUMC recently rolled out a parallel 'societal impact' evaluation to measure knowledge production, exchange and use, and economic impact across all departments. Professor Klasen commented that two essentials for annual evaluation are a good IT platform and clear indicators that draw on existing databases where possible. LUMC's indicators are agreed nationally, by the Federation of Dutch medical schools and government advisory bodies.

Integrating evaluation within Wellcome Trust

The Wellcome Trust is the UK's largest biomedical research charity, disbursing over £500m each year. Dr Liz Allen gave an overview of Wellcome's efforts to integrate evaluation approaches across funding mechanisms to better understand impact and improve funding effectiveness. Wellcome uses 10 key indicators of progress to support strategic aims such as advancing knowledge, developing people and influencing policy and practice. It tracks annual progress on these indicators using both qualitative and quantitative data. Bibliometric data provides useful additional insight into knowledge generation and impact within the broader research community, and Wellcome is experimenting with new techniques for analysis across scientific disciplines. To assess impact on training, development and research capacity, an online panel-based survey has been introduced which allows the career paths of those supported to be tracked over time. Wellcome also uses research narratives to ensure that 'proving' impact does not overshadow 'understanding' impact, compiling annual Research Profiles to shed light on progress or breakthroughs, and the role of different funders. The Trust is also building foundations for more insightful future evaluation, including work with the Research Information Network (RIN) and other funders to develop standardised citation guidelines to improve bibliometric databases. Wellcome has also been exploring opportunities to generate benchmarks between funders – although Dr Allen warned that while collaboration is valuable, each organisation has its own distinct needs.

Meta-evaluation: Ex-post review of FP6

The European Community's Sixth Framework Programmes for research and technological development (FP6) ran from 2002–2006 and had a budget of €19,235 billion, the largest multinational research programme in the world. In 2008 the European Commission set up an external panel of experts from 11 different countries to conduct an ex post evaluation of FP6 design, implementation and achievements. The panel reviewed a large evidence base of internal and external studies, including networking patterns across research areas and 'behavioural additionality' work to explore whether EC funding changed research activity. The approach worked well as a means of condensing otherwise dispersed evaluation findings. The panel recommended that a clear evaluation intervention logic should be established for future FPs, with a hierarchy of measurable objectives at different levels. Further retrospective, long-term impact studies should also be undertaken. Dr Peter Fisch of the Directorate General Research discussed the practical challenges facing evaluators in a complex, multinational environment like the EC. With diverse research topics from nuclear fusion to migration, very different objectives must somehow be accommodated within a single evaluation strategy. Differing economic and social contexts and priorities may also mean that different approaches are necessary to measure the impact of the same FP research. The lesson for evaluators: standardise where possible, but don't overdo it – diversity stimulates creativity. The challenge is to turn 'constructive chaos' into workable operational structures. A European Evaluation Network of experts from 30+ countries now provides a forum for mutual learning. ●

Impact not Ordeal: Minimising the burden on researchers

"The evaluation rules and criteria need to be clear to researchers – are we competing in the high jump or the long jump?"
[Britt Skogseid, Uppsala University]

Individual researchers and project teams are a critical data source for evaluators. However, while researchers understand that their work must be assessed, they are wary of increasingly ambitious, potentially time-consuming evaluation requirements. The feedback from Professors Sirpa Jalkanen, Åke Lernmark and Britt Skogseid was unanimous – make the evaluation process simple and consistent, allocate funding based on relevant peer review and continue to support the 'best' science.

Keep it simple

In the course of their career, a scientist will complete hundreds of funding applications and progress reports. The plea to evaluators: minimise the load through simple, well-structured processes; make criteria clear and transparent; and provide focused feedback plus practical suggestions for improvement.

While researchers are willing to be assessed on broader economic and societal impact, requirements in these less familiar areas must be clearly explained: "We should know what it is possible to show". The benefits should also be highlighted. Professor Eduard Klasen of Leiden University Medical School suggested: "Don't just give researchers directives, show them the results – how these aspects can help funding institutions secure more grants by demonstrating value to the outside world."

Provide consistency

Åke Lernmark described the strain of being evaluated by multiple institutions with different evaluation criteria, or that regularly change the criteria. Greater consistency between (and within) institutions would make life easier for researchers and reviewers, and help funders make comparisons.

The importance of peer review

Researchers want to be assessed by respected peers who understand their field. Experienced reviewers can provide constructive guidance to young investigators, and also assess more subtle factors such as teamwork and collaboration. Dr Toni Scarpa, Director of the US Center for Scientific Review (CSR) pointed out that the National Institutes of Health (NIH) distributes an extramural budget of over US $32 billion entirely on the basis of peer review, rather than 'targeting' research areas. The logistics can be challenging: CSR must assess 115,000 grant applications in 2009, necessitating 30,000 reviewers and 3,500 review meetings.

Support the 'best' science

Researchers fear that growing emphasis on accountability and financial management may stifle innovation and see good projects dropped prematurely if their impact is not immediately obvious. Sirpa Jalkanen suggested that – applying Darwinian principles – those who survive may not be the 'best' scientists but those best able to adapt to new short-term funding regimes and evaluation requirements. Longer-term projects with future potential must continue to be supported; the onus is on evaluators to identify early indicators of longer-term impact and recognise the part played by others (such as start-ups) in the impact chain, so the right balance can be struck.

Researchers urged against over-reliance on citation-based measures, which they feel may disadvantage smaller institutions or cohorts doing innovative work in niche areas and reward 'butterflies' who achieve wide publication based on 'shallow work and trendy methodologies'. A consistent record and replication by others are considered to be the most important indicators.

'Crude but crafty': getting the wider view of research impacts

Funders and researchers both want clear measures of research success, but have different definitions of 'success'. Funders need to make allocation decisions across many areas which may be at different stages of development, and show a steady stream of results. Researchers are typically focused on individual projects and future funding. While end of grant reports and case studies are an important source of information, their narrative structure makes it hard for funders to extract data and make comparisons across projects.

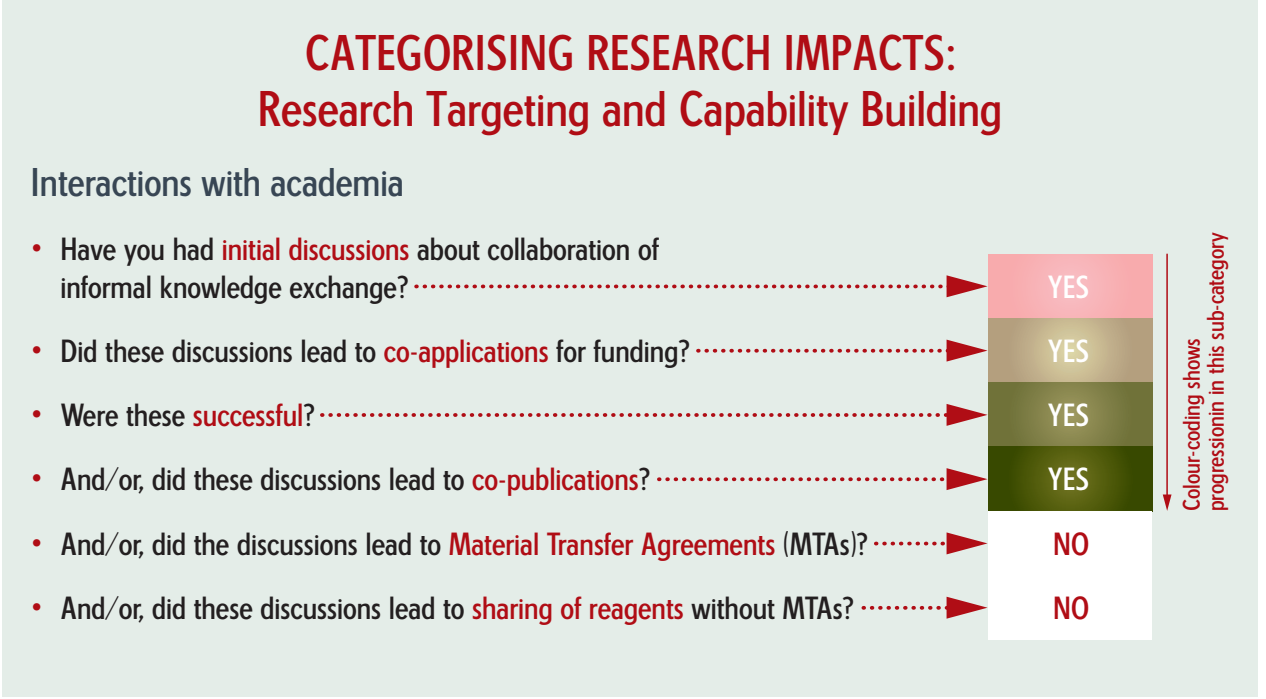
Steven Wooding presented two approaches developed by RAND Europe for the UK Arthritis Research Campaign (arc) to help map research impacts across a portfolio of work while minimising the burden on researchers. The first technique, consensus scoring, helps research teams to quantify the success of projects on a few key dimensions, thus reducing complexity. The second, a structured questionnaire, embraces complexity by seeking data on a wide range of possible impacts in a standardised way.

Collapsing complexity: Consensus scoring

The consensus scoring system presented is a 4-step process built around the five impact categories of the Payback model. Step 1: Case studies are developed using archive material, interviews, review of published materials and bibliometrics. These are reviewed and approved by the research team. Step 2: Evaluation team members score the project 1-9 on each Payback category. The scores are codified and circulated, without individual attribution. Step 3: The team discusses the scores, focusing on areas of disagreement. Step 4: The scoring exercise is repeated and usually shows greater consensus, making the scores more reliable. These final scores can be compared across projects to show impact in each Payback category. The scores can also yield insights on other dimensions such as funding mode, length of grant or impact of peer review.

Embracing complexity: The RAND/ARC Impact Scoring System (RAISS) Questionnaire

An interactive, web-based questionnaire was developed in consultation with arc senior management and over 40 arc researchers. The questionnaire – also based on Payback categories – is designed to collect detailed end-of-grant information on different types of impact without imposing an excessive time burden on researchers. The questionnaire is long (187 questions) to allow a wide range of impacts to be captured, but easy to complete as each question requires only yes, no or not known answers from the researcher. In the pilot, over 60% of researchers completed it in 30–60 minutes. The graphic below shows questions from the Categorising Research Impacts section of the questionnaire. By colour-coding answers to each question, a visual 'impact array' can be created to show strengths and weaknesses across the portfolio. Some caveats: data quality relies on the honesty of researchers, who also judge what impacts are 'significantly' attributable to the grant in question. Key benefits: the questionnaire provides a quick comprehensive overview of impacts, highlighting areas to explore in qualitative research. Because the questionnaire is quick to complete, the exercise can be repeated easily at a later date for a fuller picture of how impacts develop over time. •



"And then a Miracle happens": Increasing the impact of evaluation

"We need a 'theory of change' to clarify the links between research outputs and external outcomes and impacts."
[Dr Anas El Turabi, UK National Institute for Health Research]

For the cost and effort of evaluation to be justified, evidence must demonstrably improve research decisions and other interventions. Evaluators need to understand innovation processes, identify where and how to direct evidence and develop actionable messages for key stakeholders.

Understanding innovation processes

There is growing recognition that innovation is not a linear progression from inputs to outcomes, but arises through complex connections between different players in a broader 'innovation system'. Learning occurs across the system as players generate, test and share approaches, often tacitly through interaction, rather than formally through procedures or training. Professor Susan Cozzens of TPAC[3] outlined emerging thinking on a health innovation system spanning operational entities (e.g. health systems, hospitals, health workers), knowledge organisations (e.g. universities, research entities, information services) and governance bodies (e.g. legislators, regulators, insurers). Evaluators need to understand how different groups interact and at what level (national, regional, sectoral, local), and how ideas emerge. Another challenge is maintaining independence while working closely with those being evaluated – the ways that evaluators and researchers interact will also shape the innovation system.

Respecting political timeframes and priorities

It may be 10-20 years before the impacts of research can be fully evaluated. Meanwhile, governments – working to shorter timeframes – must decide how to allocate funds between different programmes and priorities. Dr David Cox stressed the need for indicators that show early effects of spending decisions, though a deep understanding of the innovation process is required to be certain that a change in the indicator will lead to the predicted outcome. Several speakers highlighted the need to link evaluation timetables to political timetables so policymakers receive insights in good time for the next policy intervention. Dr Peter Fisch commented that EC policy-shaping process is lengthy and involves many stakeholders, so priorities evolve. At framework programme level, there is now a commitment to timely mid-term evaluation of FP7 so the findings can inform development of FP8.

Approaches to priority-setting vary. Some governments take a hands-off approach; Dr Toni Scarpa emphasised separation of roles in the US, where Congress 'seldom' establishes biomedical research priorities – this is the role of NIH[4]. By contrast, the present Swedish government has defined 24 priority areas for science research, and will wish to assess performance against these. Former Swedish deputy minister Kerstin Eliasson stressed that research evaluation delays and constraints need to be explained to politicians seeking immediate solutions. A 'science of science policy' model could provide a framework for longer-term investment. In the US, a Science of Science & Innovation Policy Program (SciSIP) seeks to advance the basis for science policy decisionmaking.

Getting a message across

It is essential to translate complex monitoring and evaluation findings into concise, actionable messages for policymakers. Dr Peter Fisch commented that 300 page reports do not get read; recommendations from a meta-evaluation of the EC's entire research programme were summarised in a 28 page report which also laid out a future vision. From a WHO perspective, Robert Terry suggested that evaluators should combine evidence with anecdotes and examples and find fresh ways to present information. He also stressed that research evaluation is one of many inputs to health policy decisions. Providing basic systems or addressing social determinants of health may have more impact on health outcomes than further research – though these decisions should also be evidence-based. By providing simple headline figures on the economic payback of health research, the UK 'What's It Worth?' report (see p 19) has had impact with policymakers, e.g. being cited in a House of Lords debate.

The right reporting lines

Organisation design and reporting lines shape the opportunities for evaluators to have impact. For example, the EC Evaluation unit reports to the Directorate-General for Research, increasing the scope for evaluation to inform future policy, rather than the DG for Budget (as previously). In the Netherlands, ZonMw[5] is the merger of two bodies, one previously connected to the Ministry of Science, one

3 Technology Policy and Assessment Centre, Georgia Institute of Technology
4 National Institutes for Health
5 ZonMW is the Netherlands organisation for health research and development

to the Ministry of Health. The fusion allows evaluators to provide feedback to both commissioners on connections between basic and applied research, and on implementation in health care practice.

Good connections

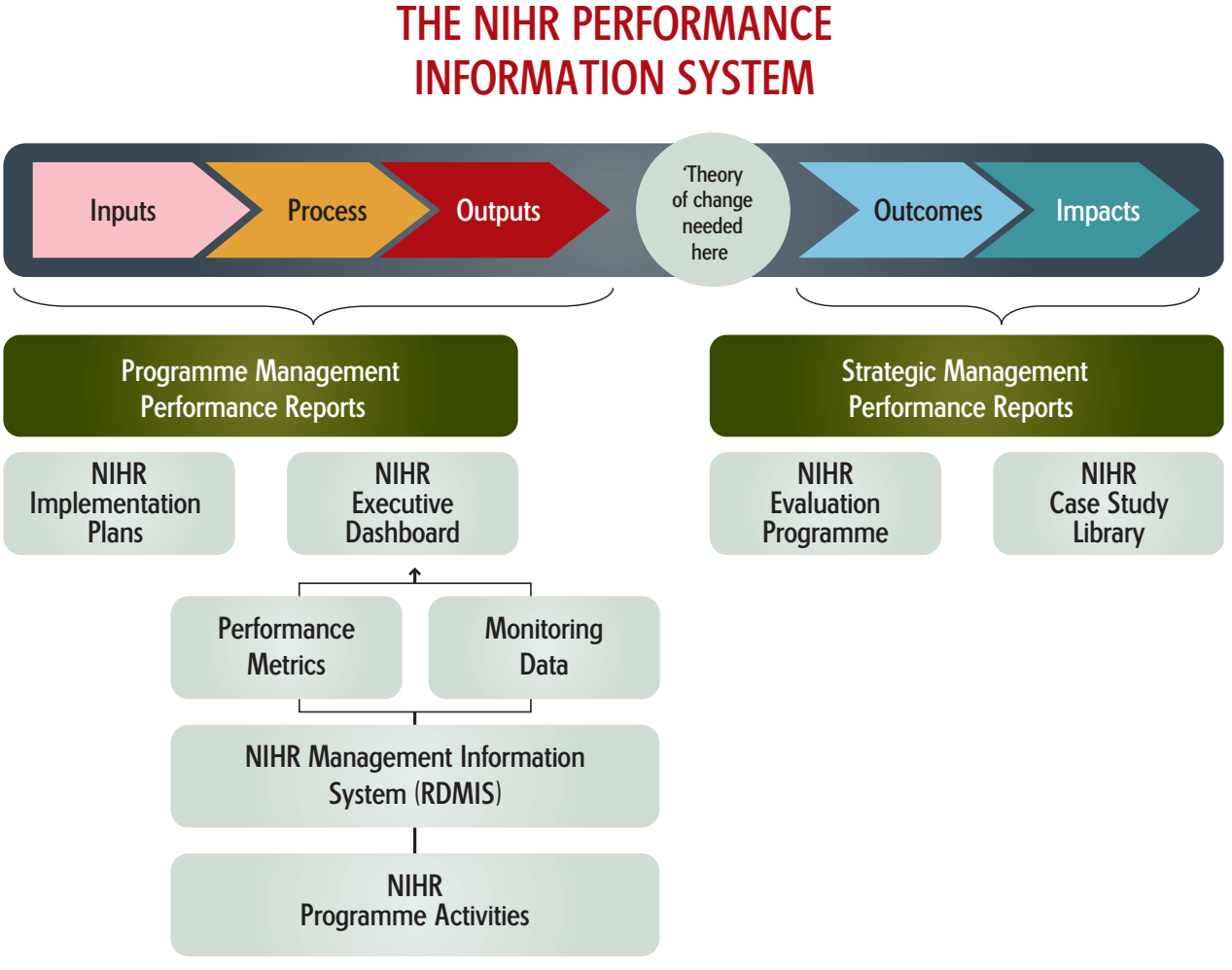
A broad reach increases the opportunity for evaluation to inform (and be informed by) other thinking. In Canada, CIHR encompasses 13 different Research Institutes, creating knowledge transfer opportunities across disciplines. LUMC [6] brings together research and patient care institutions and is co-located with 60 bio-medical companies, making it an important regional collaborator. ZonMw's board and panels include academics, patient bodies, health care professionals and policy makers to facilitate knowledge transfer and implementation of results in health care practice and policy. In the UK, NIHR [7] has been set up to strengthen systems for applied health research in the UK, and the Advisory Board involves research funders, medical schools, care delivery bodies and patients. Another new organisation, NHS Evidence (NHSE), provides an evidence base for frontline healthcare staff; key information identified by NIHR will be shared with NHSE to increase impact in the field.

Improving performance information systems

Dr Anas El Turabi described the performance information system developed by NIHR to provide programme management data for senior managers, and strategic insights to external stakeholder groups. By combining a research logic model (inputs, process, outcomes etc) with balanced scorecard performance categories (financial, internal processes, user satisfaction, learning & growth), NIHR have created an integrated approach that can track multiple indicators across the broad range of its activities. An aim, key deliverable and metric is defined for each point on the 'dashboard', using core output and outcome indicators and metrics to avoid data overload for decisionmakers. Dr El Turabi observed that a major challenge for evaluators is to use performance data to test and refine a research organisation's 'theory of change' to create one that bridges the gap between what an organisation produces (outputs) and the change the organisation wants to see in the world (impacts). •

6 Leiden University Medical Centre

7 National Institute for Health Research



Source: Department of Health UK, National Institute for Health Research

The Challenges Ahead: Issues and opportunities

"There is a lot we do not know, and a lot of issues that are difficult."

[Håkan Billig, the Swedish Research Council]

There is no simple formula for deciding 'what' science research should be funded. The future is too uncertain for anyone to identify the next scientific breakthrough, or predict the likely cost versus impact. Evaluators can, however, provide a practical evidence base on the 'how': which funding, research and knowledge translation approaches are most likely to be successful and cost-effective. This is not a simple task and the conference identified a number of issues – both methodological and policy – for future agendas.

The role of serendipity

Many scientific discoveries have come about wholly or partially as a result of serendipity: fortuitous or unanticipated discoveries or connections which lead to useful outcomes. For evaluators, the challenge is how to monitor or assess unplanned effects – for example, if a research project or programme fails to achieve its stated objective but produces other beneficial outcomes, is this categorised as a success or a failure? If a project 'fails' but develops a valuable methodology, will this be recognised in the evaluation?

Addressing time lags

The time frame for research impacts to fully emerge is often considerable. One estimate puts it at 17 years, a timeframe far longer than that of most policymakers. Evaluators need ways to show funders that their decisions are having impact while acknowledging the role of serendipity. The long-term, non-linear nature of biomedical research also makes it vulnerable to political upheaval. Evaluators and funding agencies can play a role in helping decisionmakers define appropriate longer-term (e.g. 10 year plus) strategies for research investment. To do this, evaluators will need to identify or develop 'leading' indicators that show a strong predictive power for judging the likelihood of longer term beneficial outcomes.

Attribution or contribution

It is highly probable that scientists in multiple countries are working on an area at the same time. US funders invested heavily in stomach ulcer research, but Australian scientists made the breakthrough on H.pylori. For funders seeking evidence on the performance of their investment, the challenge is how to trace, link and weight diverse contributions to outputs which may have been attributed to one team and funder. With pressure on institutions to demonstrate success to maintain their funding, some institutions take more aggressive positions than others in claiming full 'credit' – for example, where researchers receive new funding or move elsewhere, the new funder or institution may claim attribution, irrespective of other contributions. A more consistent and collaborative approach by funders would provide a more accurate picture.

Standardisation versus diversity

The evaluator's toolbox must contain a broad assortment of methodological tools that provide diverse evidence from a range of sources. However, a plethora of evaluation frameworks and classification systems prevents meaningful comparison across different projects, funders or national programmes to identify the factors that drive performance. For example, there is no standard bibliometric research classification, even though this is considered an 'easy' metric. While the term 'standardisation' is unpopular with evaluators seeking to tailor their approaches precisely to research goals and funder needs, some attempts to standardize approaches are underway – for example, the HRCS research classification system (see p 10), and the CAHS library of validated indicators and metrics (see p 18). At the same time, diversity stimulates innovation – a balance needs to be struck.

Examining the counterfactual and the halo effect

'If you think research is expensive, try disease!' [8] is a stirring catchphrase, but we should not assume that even beneficial research is better than the alternative(s). Funders and researchers prefer to celebrate successes than look for caveats, but two important questions for evaluators to consider are: 'what might have happened had research funds not been spent?' and 'are there any possible negative effects arising from this research?'. These questions are rarely asked but vital if outcomes are to be fully assessed.

Funding evaluation

Evaluation takes planning, time and money, all of which take resources from already-scarce research budgets. There is considerable debate about the 'right' amount that funders should spend on evaluation as a proportion of total budget.

[8] Mary Woodard Lasker

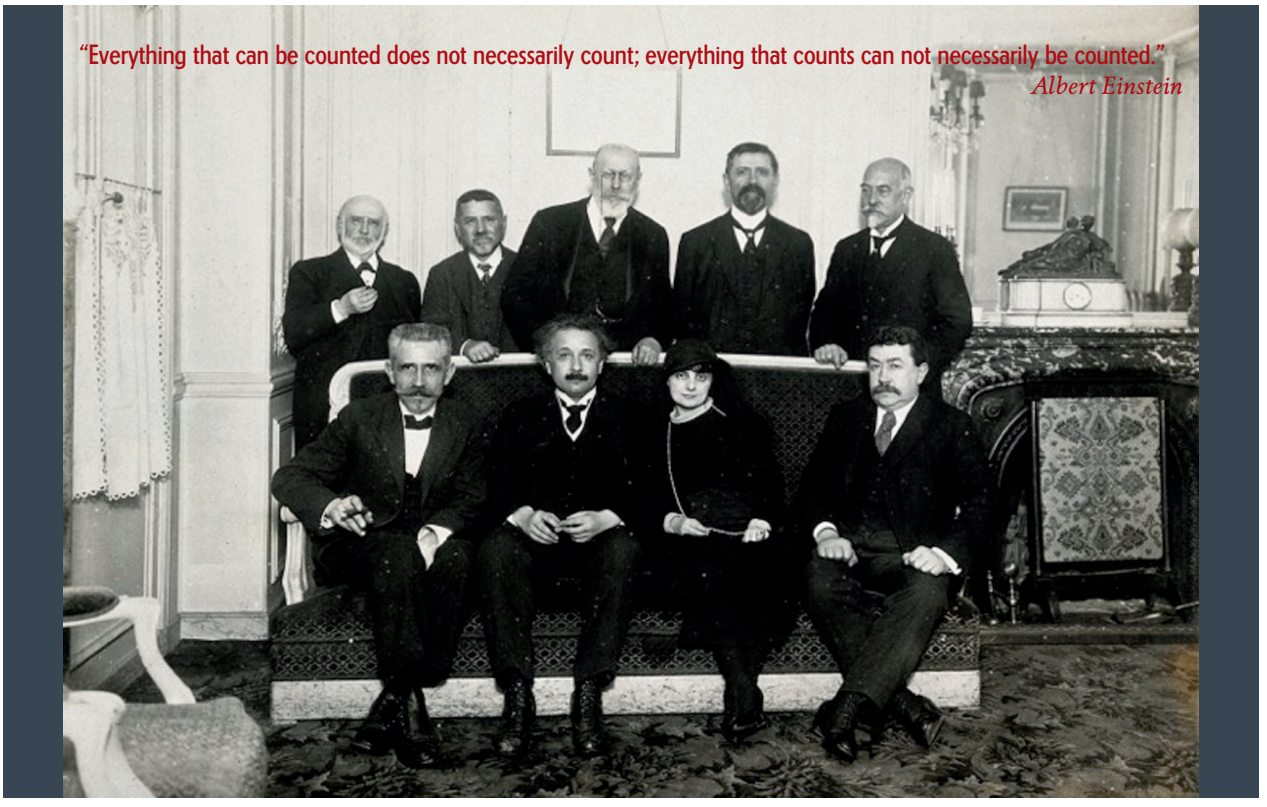
The CAHS review found a range of 0–4%, but there is currently no basis to recommend a set level, e.g. 1%. However, research evaluation is also an emerging science discipline in its own right – the 'science of science policy' – and needs further investment in methodological development if it is to have real impact on decision-making.

Making economic choices about research

For smaller or poorer nations, it is legitimate for policymakers to ask: should we do more with what we already know, i.e. focus on application and knowledge transfer to improve health outcomes quickly instead of funding new research? There is also growing recognition that synergistic collaboration between funders will maximise the value of funds spent versus duplication of efforts. The UK's 'What's it worth?' research (see p 19) suggested that while there are clear GDP benefits from having a substantial research presence, it may be economically rational for smaller nations to be 'free riders' on the basic research of others if their national science base is too small to support world-class performance in all areas. More work is needed into the efficiency of the scientific process to understand how and when broader innovation system effects start to generate benefits. Systematic analysis of economic impacts across a range of countries would also provide useful insights.

Moving to experiments

Scientific research is all about experiment, yet we rarely experiment with the way science research is funded. Allocation decisions reflect received wisdom in the form of past funding practice, peer review and past evaluation findings. Experiments might include randomly allocating research grants or funding multiple research approaches to an issue, to see whether conventional expectations of success are justified. At the margin, if a set number of grant applications are funded there is very little to differentiate the next-ranked application. It may be that sub-optimal decisions are being made, but evaluators do not yet have the tools to judge these judgements. Experiments may highlight additional success factors which are not currently recognised. •



Source: Wellcome Library, London

MAKING AN IMPACT: A PREFERRED FRAMEWORK & MENU OF INDICATORS

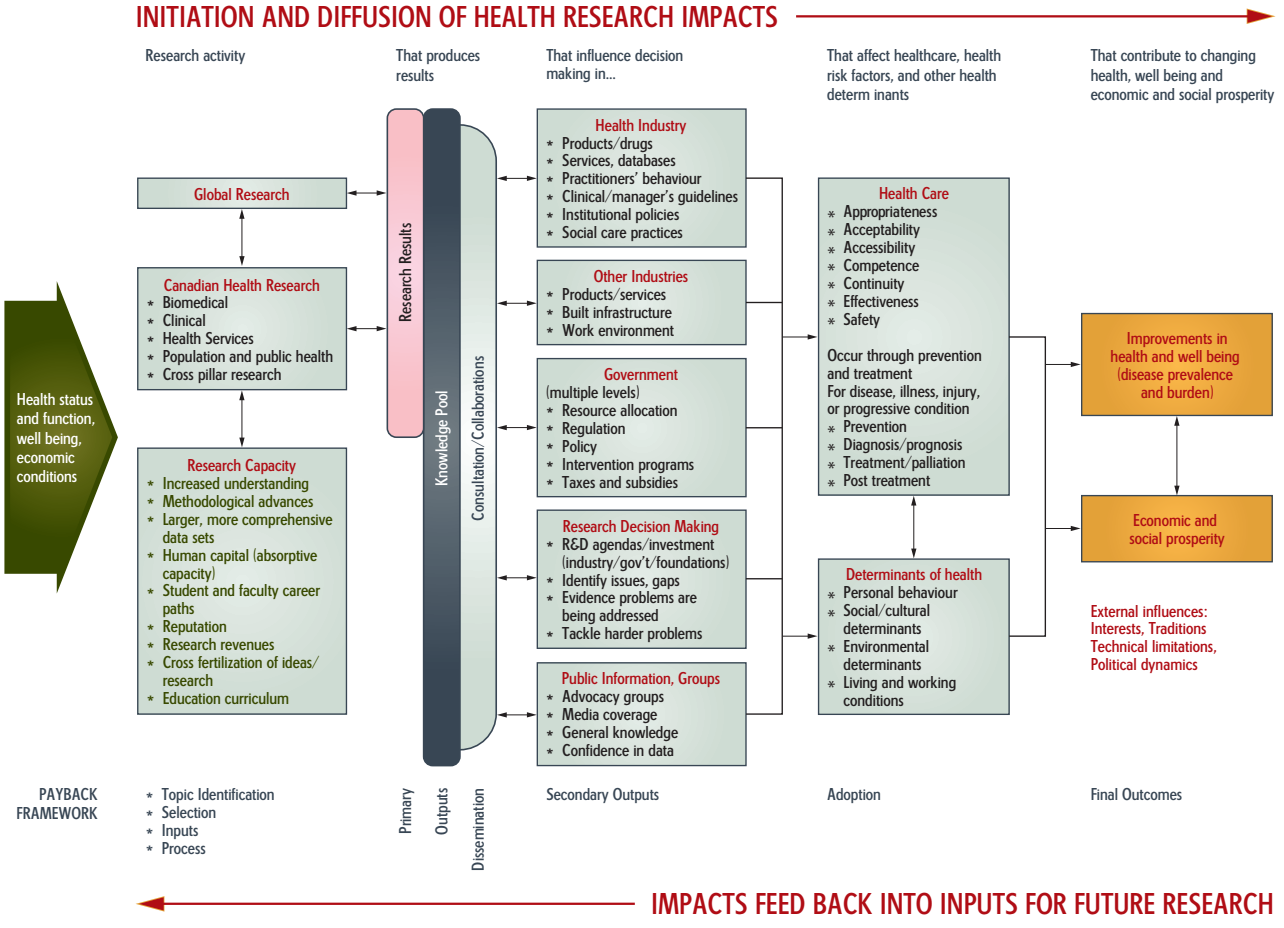
Comparability is a recurring theme in research evaluation discussions. With support from 23 sponsors and an international panel of experts, the Canadian Academy of Health Sciences tackled the question at the heart of the debate: is there a 'best' way to evaluate the impacts of health research? Professor Cy Frank introduced the results of the two-year project: a proposed new impacts framework and a starter menu of 66 validated indicators and metrics. The hope is that these will provide a common approach for all funders of health research in Canada, and – ideally – stimulate greater international collaboration.

A comprehensive, flexible and affordable framework

The CAHS team had an ambitious brief: a framework comprehensive and flexible enough to allow any funder to capture impacts in any health research area, at any assessment level from individual to international. It had to help funders (eventually) quantify their return on investment, while being practical and affordable to use. The resulting CAHS Impact framework builds on the Payback logic model to create a 'systems' approach that captures direct and indirect impacts wherever they occur. It lays out a roadmap for users (see graphic), starting with the area of research activity. Research outputs inform decision making, which leads to changes in health and in economic and social prosperity. Research impacts also feed back upstream (right to left), influencing other impacts and research.

First pick your question...

The starting point for any user is to define tightly-focused evaluation objective(s). This will help them identify where impacts may occur, and select the set of indicators and metrics that best match their needs. A provincial funder asking: "Are we building research capacity?" first needs to clarify 'research capacity' – this could include direct impacts such as quality of researchers and range of research areas represented, or indirect impacts on local decision-making. The diversity of potential evaluation questions means that the framework cannot be prescriptive in suggesting questions, only guide people to likely areas of impact, and to some tested metrics. Guidelines on research budgets and menus of priority questions and metrics are tantalising goals, but would need substantial further research. ●



Source: Canadian Academy of Health Sciences

ESTIMATING THE ECONOMIC BENEFITS OF MEDICAL RESEARCH

Few would argue with the moral case for investing in research to improve health – but what is the economic benefit to a country from doing so? In 2007, the UK Evaluation forum* commissioned a study to answer this question convincingly, by quantifying in detail the impacts of research in two areas: cardiovascular research (CVD) and mental health. The study concluded that the combined health and GDP gains from cardiovascular research provide a return of 39% to the UK economy – in perpetuity – on every £1 invested. For mental health, the figure was 37%. Professor Martin Buxton described the 5-step methodology developed by the team**, and some new questions that arise from the work.

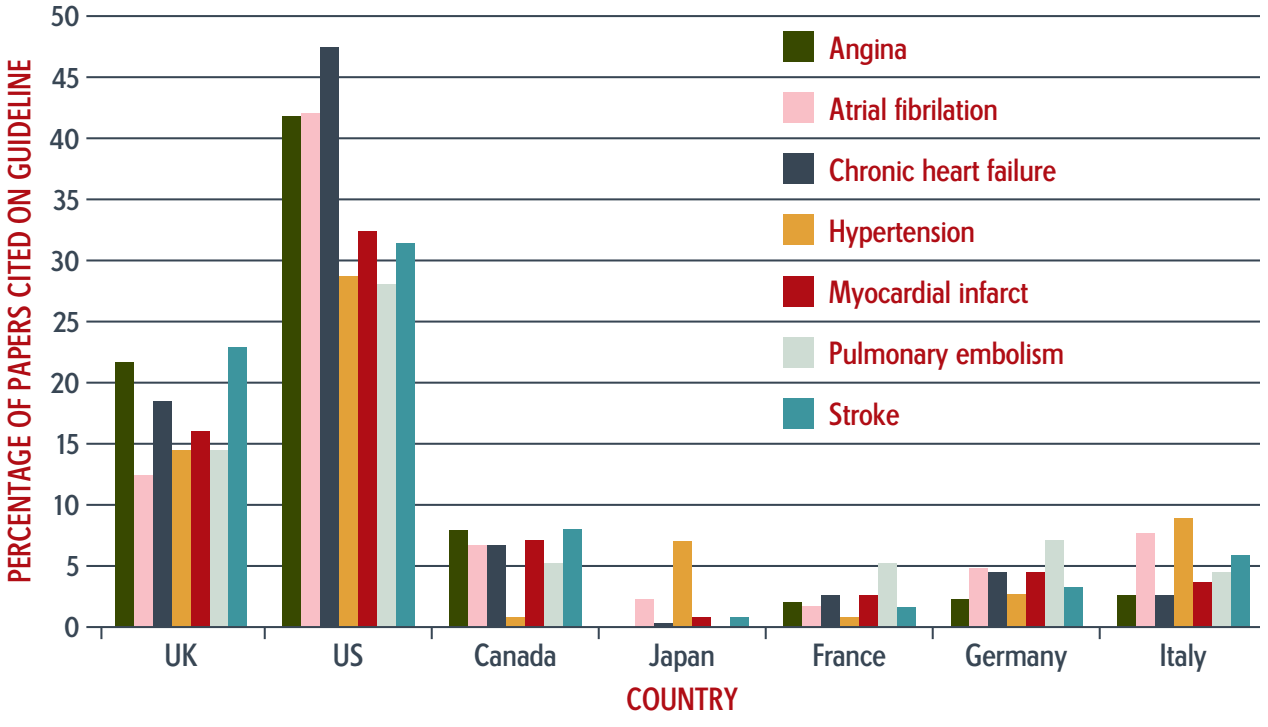
Methodology: questions, answers and some heroic assumptions

The team first identified total public and charitable expenditure on research in the chosen area – a challenging task given different research classification systems. Next, they estimated all GDP impacts for medical and non-medical sectors. Direct health gains were calculated separately, using a 'bottom up' approach to estimate the value of specific research-based interventions in additional QALYs (Quality Adjusted Life Years), minus the costs of care delivery. The fourth step addressed time lag to impact – estimated at 17 years – and the proportion of gains directly attributable to UK research (graphic). The team estimated this at 17%, a mid-point in a wide range (12–23%) reflecting the global nature of research and the difficulties of attribution. Finally, the team calculated overall returns for optimistic, pessimistic and mid-case scenarios.

Is research a 'public good'?

The study highlights the need for standardised research classification to support comparable work in other disease areas and to analyse global research impacts. It also raises fascinating science policy questions about the importance of local research in achieving local health and GDP gains. Could a nation choose to be a 'free-rider' on research funded elsewhere? Or is local research capability essential for efficient adoption of new ideas? More work is needed on how well and rapidly different nations, with different research spends, adopt valuable new technologies. Meanwhile, finding ways to shorten the time lag to impact would significantly improve all research returns. ●

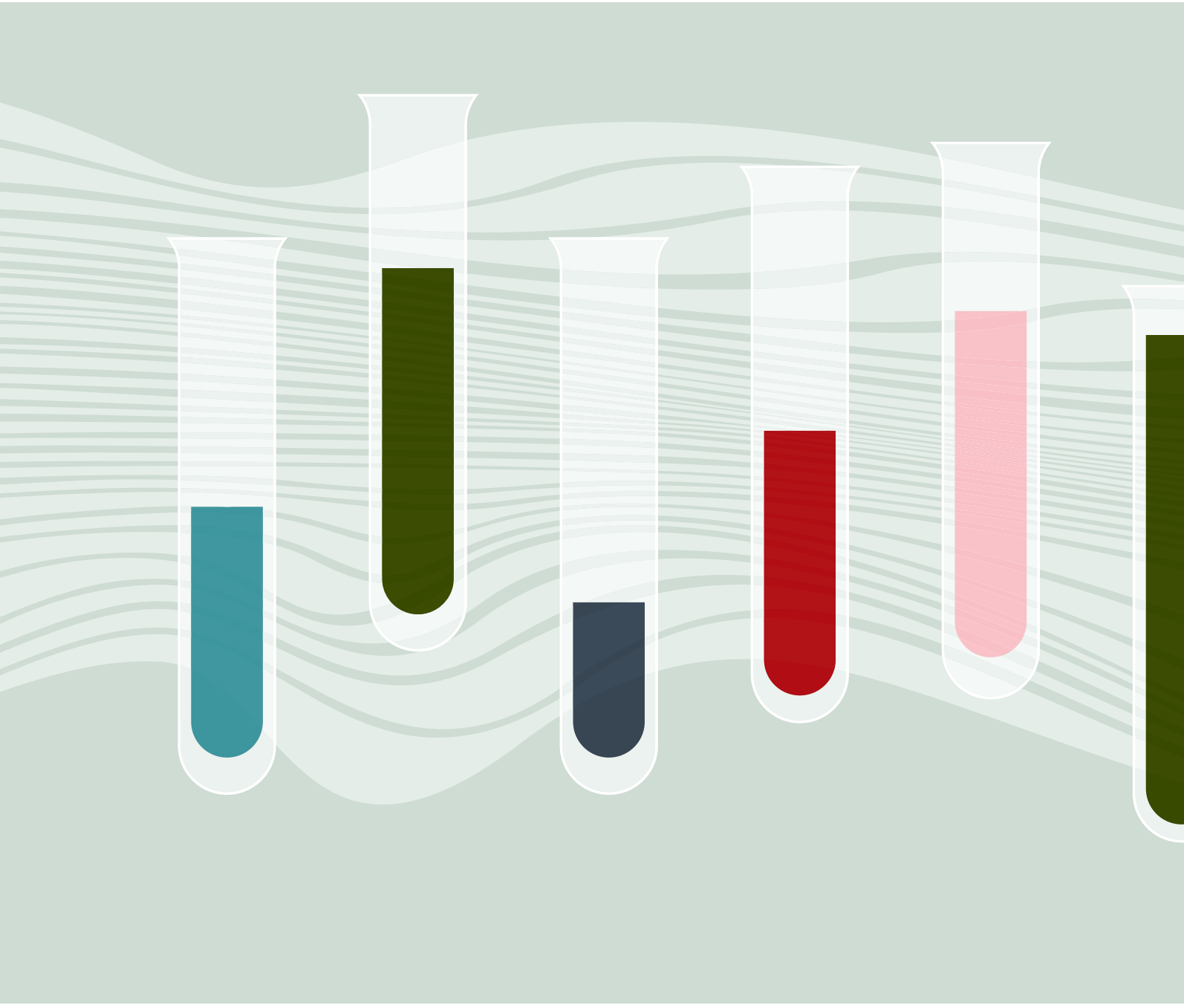
NATIONALITY OF PAPERS CITED IN SEVEN CARDIOVASCULAR CLINICAL GUIDELINES



Published between 2003 and 2007

* Specifically, the UK Medical Research Council, Wellcome Trust and the Academy of Medical Sciences

** The joint research team was HERG (Health Economics Research Group, Brunel University), RAND Europe and the Office of Health Economics.



The Swedish Research Council is a government agency that provides funding for basic research of the highest scientific quality in all disciplinary domains. Besides research funding, the agency works with strategy, analysis, and research communication. The objective is for Sweden to be a leading research nation.