UNESCO Sponsored Mission to Mongolia

October 29th - November 7th 2005

Consultant's Report

Consultant:

Professor Tim Turpin
Acting Director, Australian expert Group in Industry Studies (AEGIS)
University of Western Sydney

Australian Expert Group in Industry Studies, UWS
Postal Address: PO Box Q 1287 QVB Post Office NSW 1230
Location Address: 8th Floor, 263 Clarence Street Sydney
Telephone: 61 2 8255 6200 Facsimile: 61 2 8255 6222

CONTENTS

UNESCO MISSION TO MONGOLIA OCTOBER 29TH – NOVEMBER 7TH

MISSION REPORT

Prepared by Professor Tim Turpin

Preparation for the Mission

Professor Tim Turpin was appointed as consultant to UNESCO Beijing Office to participate in the National Seminar for Science and Technology focused on 'science and technology sectoral analysis reviews' scheduled in Ulaanbaatar, Mongolia October 31st and November 1st. The consultant was also asked to present a public lecture on World Science Day in Ulaanbaatar at the Mongolian University for Science and Technology

Preparation for the mission included:

- reviewing the previous Draft Master Plan and identifying areas for further discussion and input during the mission;
- preparing a presentation for the national seminar; and,
- preparation a presentation for a public lecture on World Science Day.

These activities were completed during the week prior to the Mongolian mission.

Professor Turpin travelled to Ulaanbaatar from Sydney via Seoul on October 28th and arrived in Ulaanbaatar 29th October.

Discussion on progress through 2005

On Sunday 30th October Professor Turpin met with the team from the Department of Science and Technology to review progress and discuss inputs to the national Seminar.

The S&T policy team had completed an impressive amount of work through the year. This work included: national level data collection and analysis and comparisons with other countries; case-study data collection and analysis focused on industry; and sectoral analysis of higher education and industry collaboration.

Two issues with implications for further work toward the national S&T strategy stood out. **The first concerned the issue of priority setting**. There is clearly considerable scientific capacity within the Mongolian higher education and public research institute sectors. A major task for future national development is to draw on and utilise this capacity for the future development of the knowledge base that must underpin socio-economic development through, at least, the next two decades. This will require **not simply** establishing a set of scientific or disciplinary priorities, **but rather,** establishing an on-going processes for directing public research investments toward solving key problems and bottle necks in the national development strategy.

As we have noted in reports from pervious missions it is a process of engagement and dialogue between government, scientists, industry and society generally that is needed to ensure continuity between the production, capture, transmission and application of knowledge. The National Seminar was an important step toward this process. *As discussed below some further work on 'priority setting' is proposed in order to enhance the capacity for government to maximise socio-economic benefit, across all sectors, from long-term research investment.*

The second issue concerned the monitoring, evaluation and benchmarking of research investments and outcomes. Effective monitoring and establishing *appropriate* indicators for benchmarking progress toward S&T goals is an essential component of long term strategic planning. The S&T policy team had made considerable progress toward this task. However, in the absence of long-term national strategic goals for steering public research investments it is difficult to identify and construct appropriate indicators to monitor progress toward such goals and to inform the policy process that delivers research investment.

The consultant focused in particular on this issue throughout the mission drawing attention to the experiences of other small nations in establishing research priorities and consequently the need for indicators to assess the outcome and impact of investment initiatives directed toward these priorities.

In the Mongolian context is important to recognise that the process of socio-economic transition is still very much underway. In order to derive national benefit from science for development the science *system* must undergo parallel reforms. This is already well underway but still lagging behind the broader process of transition. There is an urgent need for a broad community understanding that **priority setting does not mean certain fields of science become more important than others**. Rather, it means identifying ways that **all fields of science can contribute toward the knowledge base necessary for maximising progress toward the broader national goals for socio-economic progress.**

The National seminar served well as a forum for discussing this issue and moving toward that goal.

Participation in the National Seminar

The National Seminar was held over two days: October 31st and November 1st. The first day provided for input from speakers from universities, research institutes, government agencies and UNESCO. The focus of the presentations was on current progress toward implementing an S&T strategic in Mongolia.

Professor Turpin's presentation was during the first day of the seminar. A copy of the power point presentation is included in this report as Attachment 1.

The second day of the seminar focused mainly on progress toward implementation of the strategic plan and mechanisms for maximising consensus during the process. Four workshop discussion groups worked well toward this objective.

After the workshop the consultant discussed the workshop outcomes with the Deputy Director of the Department of Science and Technology, MECS. Meanwhile the Department is reviewing all of the working group findings. They will draw on these findings for subsequent work toward the national strategy.

Public lecture on World Science Day.

This was hosted by the Mongolian University for Science and Technology on November 3rd. Dr Aoshima, Director UNESCO Beijing chaired the session. The power point presentation is included as Attachment 2.

Follow-up Activities

Following the national seminar and subsequent discussions with Ministry Officials and UNESCO it was agreed that some further inputs to the S&T policy making process were necessary in order to achieve the implementation deadline of June 2006.

As noted above there are two areas where further work should be targeted: priority setting and evaluation. There a range of tools appropriate for both these area that have already served many small countries well in bringing their science capacity to bear on national objectives. It is recommended here that an activity should be planned early in 2006 in order to adapt and transfer some of these tools to the Mongolian context via the Ministry of Education, Science and Culture.

Priority Setting

As noted above this is perhaps the single most important issue for Mongolian Science S&T policy. International experiences have shown that the key to successful priority setting is not so much about setting specific scientific targets but rather establishing an on-going process that draws science (through scientists) into on-going debates and decisions about *how* science can contribute to a wide range of pressing national socio-economic objectives. Useful methodological tools used elsewhere have included: 'scenario planning'; 'foresight activities'; 'action agendas'; and 'priority setting processes'. All have different strengths and weaknesses and their useful application depends much on the national context. All serve in one way or another to build useful bridges between scientists across different sectors, industry, social planning agencies, NGOs and broader social expectations and aspirations.

From our discussion we believe that the use of an 'Action Agenda' methodology directed toward one or two key sectors in the Mongolian economy would greatly assist the Ministry in building cross-sector research collaboration. Most importantly, transferring the method to the Mongolian context through such a pilot initiative will enhance the capacity for on-going policy review, monitoring and evaluation and finetuning national S&T initiatives in the longer term.

Evaluation and monitoring

Evaluation and monitoring is an essential part of the policy process. Ideally it should be carried out in parallel with any priority setting process. The Ministry has made great steps toward establishing such a process. The process will be greatly enhanced by linking evaluation and monitoring to the S&T planning process. It is recommended here that the priority setting pilot activity proposed above should be carried out in parallel with the identification of benchmark indicators suitable for monitoring outcomes and impact in the sectors targeted for the pilot activity.

Recommendation

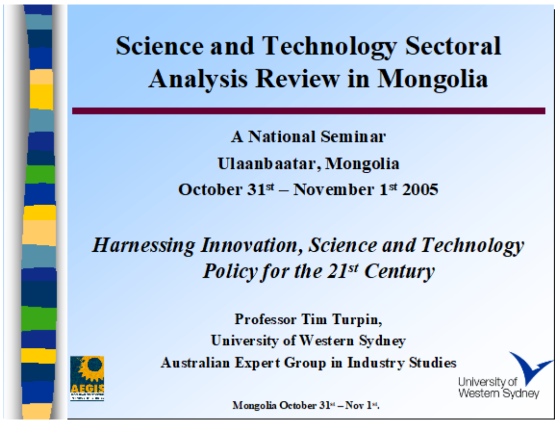
It is recommended that the Ministry of Education Culture and Science discuss options with UNESCO for carrying out a priority setting activity along the lines proposed above early in 2006.

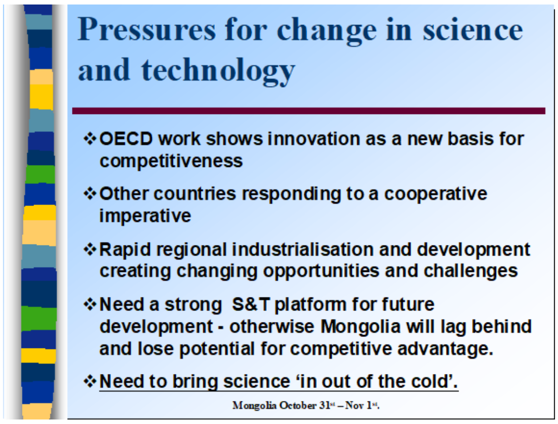
The objective of this activity should be to:

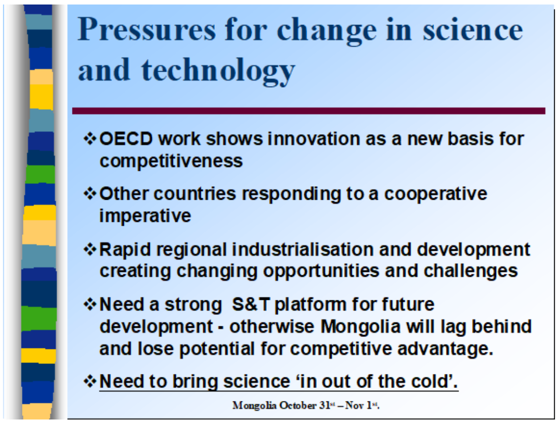
- adapt and transfer a priority setting methodology appropriate for Mongolia's current socio-economic objectives and scientific capabilities;
- build further consensus among key S&T agencies and stakeholders toward a national S&T strategy;
- enhance the capacity for developing national indicators and benchmarks for monitoring research outcomes and impact; and
- establish research priority setting processes and evaluation as a central component for implementation of the national S&T strategic plan.

Attachments

Attachment 1: S&T Workshop Presentation







International Experiences

- S&T policies for the 21st century require a *flexible system*
  - capable of responding to changing technological demands as global and local technological and business environments change
- *Coordination* is required
  - between S&T incentives and those in other sectors to enable innovation capacity to cross critical technology thresholds

International Responses

- These transition efforts involved *dual* policies:
  - measures to strengthen the support institutions (the science base)

    and

  - measures to strengthen activities in the firms, especially those to stimulate firm demand for improved technology (the innovation base)

International science policy – focus and objectives

Seek to build critical mass concentrated *around* specific fields of knowledge and *across* institutional structures.

They are systems of research 'organisation' that cross university and research institute boundaries and industry in innovative ways.

Shifting the policy perspective

- Fragmentation → Integration
- Static competition → Dynamic competition
- Discipline or sectoral priorities → Clusters of priorities
- Institution or discipline based funding → Performance based funding
- Public sector action and policy targets → Parallel public/pvt. actions and targets
- Technology development *for* industry → Technology development *by* industry
- Project or program evaluation → Systemic evaluation

Expectations from investments in research

1. Generating new knowledge
2. Solving problems
3. Enhancing scientific capacity
4. Underpinning an economic platform for development

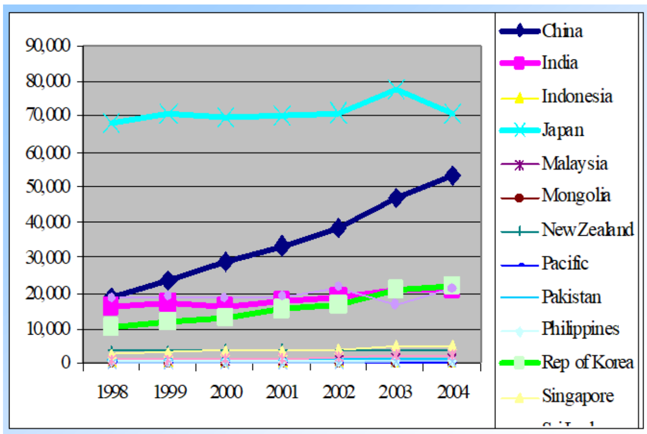
Mongolia October 31st – Nov 1st.

Trickle down effect

- These expectations and demands flow from government *and* the broader community impacting on research funding agencies, research institutes, universities and the practitioners.

Mongolia October 31st – Nov 1st.

	1998	1999	2000	2001	2002	2003	2004
China	18,833	23,398	29,004	33,206	38,469	46,900	53,235
India	16,037	17,104	15,983	17,501	18,525	20,803	20,867
Indonesia	342	381	422	486	395	489	456
Japan	68,585	70,435	69,773	70,215	71,207	78,046	70,865
Malaysia	798	869	814	922	934	1,171	1,257
Mongolia	34	38	45	39	41	56	60
New Zealand	3,793	3,798	3,847	3,842	3,793	4,135	4,046
Pacific	100	141	108	101	108	130	105
Pakistan	601	577	596	531	691	763	904
Philippines	311	344	351	315	410	440	424
Rep of Korea	10,458	11,894	13,200	15,519	16,642	20,529	22,031
Singapore	2,490	3,046	3,392	3,802	4,238	4,846	5,118
Sri Lanka	124	168	167	157	176	264	227
Thailand	935	1,043	1,185	1,331	1,591	2,048	2,049
Australia	18,302	18,898	18,486	18,968	21,601	16,573	20,985
All countries	*141,743*	*152,134*	*157,373*	*166,935*	*178,821*	*197,193*	*202,629*

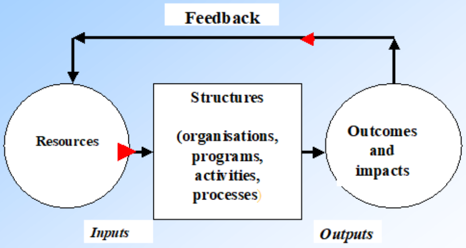


Evaluation

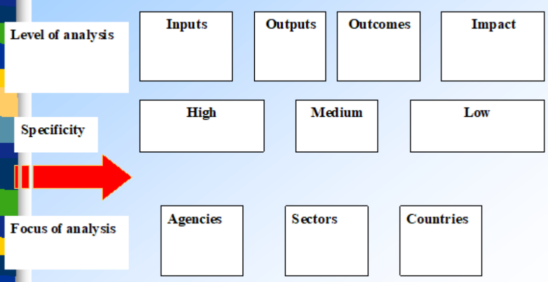
- **National Evaluation System**
  - **Linked to priority setting process**
- **Cross sector outputs and outcomes**
  - **Universities, Centres of Excellence and , Institutes**
- **Stakeholder involvement in process of**
  - **Establishing performance criteria**

Mongolia October 31st – Nov 1st.

Model of the research system for evaluation purposes



Inputs Outputs, Outcomes and Impact



Benchmarking HMRD

- **Research evaluation concerns:**
  - Results, efficiency and effectiveness
  - Demonstrating the results to particular audiences;
  - Assessing the value of the work
- **Benchmarking concerns comparisons of performance**
  - It can be qualitative or quantitative

Options for development

- Internationally, many agencies have developed formal frameworks for reporting results
- However, only limited scope for benchmarking performance between countries

Mongolia October 31st – Nov 1st.

A more structured approach

- International benchmarking
  - Integrate private and public sector to assess scope and strength of national performance
- National benchmarking
  - Agency contributions to national priorities
- Agency institute benchmarking
  - Cost, efficiency, program outputs
- Program benchmarking
  - Descriptive performance assessment

Mongolia October 31st – Nov 1st.

Building a common classification

- Input classifications
- Sector classifications
- Specified targets and beneficiaries

Mongolia October 31st – Nov 1st.

Responding to different audiences

- Evaluation that makes sense to researchers *and* other stakeholders
- Identification of *common* objectives
- Evaluation that informs research specialists (quality) and the users and practitioners (relevance)

Mongolia October 31st – Nov 1st.

Some monitoring and evaluation questions for Mongolia

- Impact in strengthening collaborative processes?
- The flow of skilled personnel across institutions and sectors?
- Capacity and outputs from research strengths and priorities?
- Progress in building critical mass in key research and teaching areas to complement major national socio-economic priorities?

Mongolia October 31st – Nov 1st.

Attachment 2: Public Lecture on Science and Technology Policy for World Science Day (MUST)

Slide 1

AUSTRALIAN EXPERT GROUP IN INDUSTRY STUDIES

University of Western Sydney

Capturing small country advantage:

a comparative framework for evaluating research outcomes

Tim Turpin
Australian Expert Group in Industry Studies
University of Western Sydney

The evaluation context

A wizard thrives in the company of believers and in the absence of formal evaluation.

Science thrives in the context of critical scepticism and evaluation

Slide 3

The Global Context

- **The 21st century presents an industrial landscape dominated by strategic networks**
- **that draw together skills and knowledge about science, technology, products, services and business systems into 'innovation networks'.**
- **Key actors in the networks include firms, research institutes, training institutes and other knowledge intensive institutions such as universities and hospitals.**
- **These networks are constantly changing as the dominant players make decisions as to whether to expand or reduce access to the networks, whether to open entry to new members or to 'pull up the drawbridge'.**
- **Innovation networks transcend national boundaries.**

The role of science in innovation networks

- Science is a key component of this industrial landscape
- The demand and potential of science to contribute to this landscape has, in many countries, has brought science '***in from the cold***'.
- Into an environment dominated by ***creativity, knowledge, technique, investment*** and ***risk***.

Slide 5

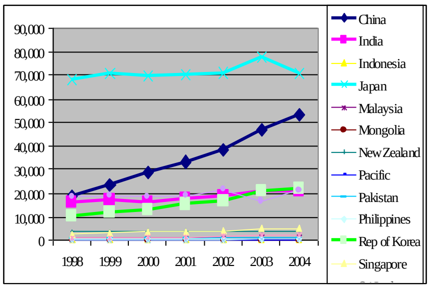
Transition to a new economic base: toward a knowledge economy

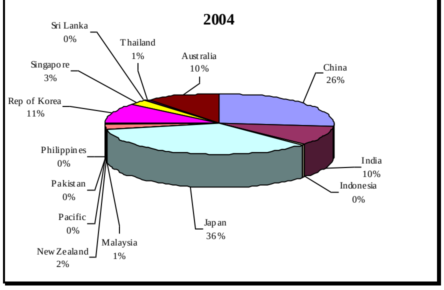
- Cooperation is critical for gaining access to human resources, new knowledge, ideas and know how.
- Dynamic rather than static competition is required
- Few countries have the capacity to be self sufficient in terms of scientific, technical and innovative capacity.

Publications in ISI English language journals

	1998	1999	2000	2001	2002	2003	2004
China	**18,833**	**23,398**	**29,004**	**33,206**	**38,469**	**46,900**	**53,235**
India	**16,037**	**17,104**	**15,983**	**17,501**	**18,525**	**20,803**	**20,867**
Indonesia	**342**	**381**	**422**	**486**	**395**	**489**	**456**
Japan	**68,585**	**70,435**	**69,773**	**70,215**	**71,207**	**78,046**	**70,865**
Malaysia	**798**	**869**	**814**	**922**	**934**	**1,171**	**1,257**
Mongolia	**34**	**38**	**45**	**39**	**41**	**56**	**60**
New Zealand	**3,793**	**3,798**	**3,847**	**3,842**	**3,793**	**4,135**	**4,046**
Pacific	**100**	**141**	**108**	**101**	**108**	**130**	**105**
Pakistan	**601**	**577**	**596**	**531**	**691**	**763**	**904**
Philippines	**311**	**344**	**351**	**315**	**410**	**440**	**424**
Rep of Korea	**10,458**	**11,894**	**13,200**	**15,519**	**16,642**	**20,529**	**22,031**
Singapore	**2,490**	**3,046**	**3,392**	**3,802**	**4,238**	**4,846**	**5,118**
Sri Lanka	**124**	**168**	**167**	**157**	**176**	**264**	**227**
Thailand	**935**	**1,043**	**1,185**	**1,331**	**1,591**	**2,048**	**2,049**
Australia	**18,302**	**18,898**	**18,486**	**18,968**	**21,601**	**16,573**	**20,985**
All countries	*141,743*	*152,134*	*157,373*	*166,935*	*178,821*	*197,193*	*202,629*

Slide 7





Slide 9

US Patents for Inventions Granted to Residents of ASEAN Countries and Australia

	1999	2000	2001	2002	2003	Total since 1963
Australia	707	705	875	859	900	14725
Brunei	0	0	0	0	0	1
Indonesia	5	6	4	7	9	151
Malaysia	30	42	39	55	50	356
Myanmar	0	0	0	0	0	4
Philippines	11	2	12	14	22	245
Singapore	144	218	296	410	427	2098
Thailand	20	15	24	44	25	205
Vietnam	2	0	0	0	0	11
Total US patents for invention granted	153486	157495	166037	167333	169028	3583814

Patents Granted in ASEAN Countries and in Australia

	1993	1998	1999	2000	2001	2002
Australia	**12728**	**14784**	**13528**	**13916**	**13983**	
Philippines	940*	565	648	566	1092*	
Singapore		2291	4410	5090	7220*	
Vietnam	16		490	727		
Thailand	451	723	392*	416*	796*	1102*
Malaysia	1121	567*+	720*+		1470*	
Indonesia	2069*	1846*	2936*	3890*	3926*	

Source: WIPO. WIPO figures are used whenever available.
* national source
\+ including utility inventions

Slide 11

Applications to Register Industrial Designs in ASEAN Countries and in Australia

	1994	1995	1996	1997	1998	1999	2000	2001
Australia	4348	4045	4415	4207	4105	4371	4255	4119
Philippines	595*	497*	839	878	726	767	819	393*
Thailand	962	904*	960	1148	1338	1721*	2697	2662*
Vietnam	716	1131	1646	1153	1057*	1036	1207	
Indonesia								1403*

Source: WIPO. WIPO figures are used whenever available.
* national source

The small country challenge

- Establishing priorities
- Tapping into and capturing global knowledge and expertise
- Building a national strategy for development across ministries and key agencies
- Monitoring progress and refining the strategy

Slide 13

Small country strategies for science and technology for development

Two key questions:

- Where are we going?
- How far have we travelled?

Brunei: establishing priorities

- Brunei International Finance Centre
- Halal Food
- ICT
- Tourism
- Transhipment
- Oil and gas related activities

Slide 15

Brunei's strategy

- **Enhancing economic growth and diversity by building a long term innovation capacity;**
- **Identifying a package of innovation programmes and investments for '*Creating Brunei's Future*';**
- **Mechanism: targeted research and innovation**
  - ***Coordination*** across portfolios to link programmes to national development strategy through input of S&T
  - ***Monitoring*** and ***evaluating*** progress through agreed innovation and S&T indicators
  - ***Building*** a flexible S&T platform by setting up teams across organisations
  - ***Building on the best*** of Brunei's science and social science capabilities.

Iceland's Strategy

- Link S&T strategy to national industrial and social priorities
- Build on geothermal power
- Build on regional strengths and opportunities
- Establish a role in a regional value chain
- Establish a system for setting S&T targets, monitoring and evaluating outcomes and assessing progress.

Slide 17

Ireland

- A transformation from dependency to dynamic competitiveness
- Built on selected priorities for development
- Establishing a key position in European value chains
- Clearly established monitoring and performance system

Government and social expectations from investment in research

1. Generating new knowledge
2. Solving problems
3. Enhancing scientific capacity
4. Underpinning an economic platform for development

Slide 19

Trickle down effect

- These expectations and demands flow from government *and* the broader community impacting on research funding agencies, research institutes, universities and the practitioners.

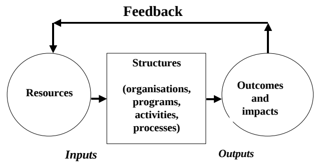
How far have we travelled?

Benchmarking HMRD

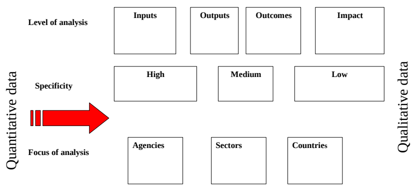
- Research evaluation concerns:
  - Results, efficiency and effectiveness
  - Demonstrating the results to particular audiences;
  - Assessing the value of the work
- Benchmarking concerns comparisons of performance
  - It can be qualitative or quantitative

Slide 21

Model of the research system for evaluation purposes



Inputs Outputs, Outcomes and Impact



Slide 23

Three problems in international benchmarking (HMRD)

1. Defining the HMRD system
   - Structures, institutions and organisations are specific to sectors and countries
2. Classifying the purpose of HMRD
   - A lack of internationally accepted functional classifications
3. Comparing between agencies, sectors and countries
   - Matching appropriate levels of specificity

International approaches to HMRD

- Similar approaches and measures as other fields
- Hierarchical approach linking top level objectives to research output level
- Integration of qualitative and quantitative data
- Focus on beneficiaries as well as standard statistical indicators

Slide 25

Measuring intensity of HRMD inputs

Country	HMRD as % of health care expenditure	Institutional HMRD funding per capita (US$)
USA	1.50	$64.06
Singapore	1.24	$8.14
New Zealand	0.54	$6.40
UK	0.53	$8.92
Canada	0.41	$8.02
France	0.39	$8.82
Australia	**0.38**	**$6.89**
Netherlands	0.35	$7.53
Germany (Est. b)	0.31	$8.43
South Africa	0.31	$0.70
Switzerland	0.30	$11.33
Germany (Est. a)	0.20	$5.50

Source: Turpin, Wixted and Garrett-Jones, 2003

Assessing functional priorities in HMRD

Health objective	Australia 2002	Germany 2002	NZ 2001	Canada 2001	UK 2001–02	USA 2002	Singapore 2000
Cancer, cancer prevention and related disorders	**14.9**	5.4				21.0	
Cardiovascular health and diseases	**11.0**	3.7				8.7	
Endocrine diseases and diabetes	**4.2**						
Injury	**3.2**		3.6				
Mental health and neurosciences	**16.0**	13.2	15.3		16.7	25.3	
Respiratory diseases	**4.9**						
Bone, joint and muscle diseases	**4.0**						
Human genetics and inherited disorders	**3.8**	14.2	17.1				
Infection and immunity	**16.4**	13.4			16.7		
Liver, kidney and gastro-intestinal health and diseases	**4.9**						
Other health issues, diseases and conditions	**10.8**						
Reproductive health	**4.3**						
Social and environmental health issues	**1.6**	1.2					
Population groups and health		0.8	6.0		14.0		
Health sector/ system management		3.9	6.2				
AIDS research				10.7		10.7	10.7

Source: Turpin, Wixted and Garrett-Jones, 2003

Slide 27

Assessing priorities in relation to burden of disease

- The ability to benchmark investments according to burden of disease is limited by a lack of common 'disease-based' classification for HMRD expenditures

Assessing outputs and impact

Research field[a]	Australia – no. of papers	Australia – no. of citations – actual	Australia – citation rate – actual	Australia – citation rate – expected	World – citation rate	Australia – actual citations/ expected	Australia – actual citations/ world[b]	Australia – share of world papers
Biological Sciences								
Biochemistry and cell biology	6,790	50,037	7.4	7.3	8.3	1.01	0.89	2.4%
Genetics	2,100	15,255	7.3	7.4	8.2	0.98	0.89	3.5%
Microbiology	1,668	8,542	5.1	5.0	5.0	1.02	1.02	2.8%
Biotechnology	2,147	10,058	4.7	4.6	4.6	1.02	1.03	2.6%
Medical and health sciences	**30,663**	**143,234**	**4.7**	**4.5**	**4.8**	**1.04**	**0.98**	**2.8%**
Immunology	2,696	20,806	7.7	7.3	6.8	1.06	1.13	2.9%
Pharmacology and pharmaceutical science	2,654	10,028	3.8	4.1	4.0	0.92	0.95	2.1%
Medical physiology	2,253	7,731	3.4	3.6	3.9	0.95	0.89	4.1%
Neurosciences	3,048	16,553	5.4	5.9	6.3	0.93	0.87	2.4%
Clinical sciences	19,547	93,902	4.8	4.5	4.7	1.06	1.03	2.7%
Public health and health services[c]	3,412	10,235	3.0	3.0	3.3	1.00	0.90	3.7%
Medicine - general and internal	2,215	12,754	5.8	5.2	5.7	1.12	1.01	3.1%
Multidisciplinary science	**1,389**	**16,748**	**12.1**	**13.3**	**11.4**	**0.91**	**1.06**	**1.8%**

Source: Data from Butler (2003)

Slide 29

Options for development

- Many agencies have developed formal frameworks for reporting results
- However, only limited scope for benchmarking performance between countries

1. A more structured approach

- International benchmarking
  - Integrate private and public sector to assess scope and strength of national performance
- National benchmarking
  - Agency contributions to national priorities
- Agency institute benchmarking
  - Cost, efficiency, program outputs
- Program benchmarking
  - Descriptive performance assessment

Slide 31

2. The need for common classifications

- Input classifications
- Sector classifications
- Specified targets and beneficiaries

3. Responding to different audiences

- Evaluation that makes sense to researchers *and* other stakeholders
- Identification of *common* objectives
- Evaluation that informs research specialists (quality) and the users and practitioners (relevance)

Slide 33

The capacity of 'wizards' to deliver great outcomes and potential wealth can be convincing *only in the absence* of definable and reliable performance indicators and collaboration.

International experience has shown that the capacity of science to deliver social and economic benefit can, in contrast, best be defined, understood and marshaled with the assistance of sound evaluation indicators and collaboration.

Attachment 3: World Science Day text of lecture presented at Mongolian University for Science and technology

Capturing Small Country Advantage:

Establishing a comparative framework for evaluating research outcomes*

Invited public lecture on World Science Day,
Ulaanbaatar, Mongolia November 3rd 2005.

Professor Tim Turpin
Australian Expert Group in Industry Studies, University of Western Sydney, Australia

* This paper is based on a recent publication

Sam Garrett-Jones, Tim Turpin and Brian Wixted. 'Some international benchmarks for evaluating Australian health and medical research', *Research Evaluation,* December 2004.

Capturing Small Country Advantage:
Establishing a comparative framework for evaluating research outcomes

A wizard thrives in the company of believers and in the absence of formal evaluation. Science thrives in the context of critical scepticism and evaluation

ABSTRACT

Governments around the world have been facing increasing demands for greater accountability and efficiency in their public investments in research. In response many governments have turned to international benchmarking as a means to establish comparisons for monitoring and assessing performance across a diverse range of scientific disciplines and funding mechanisms. Health and medical research, until quite recently somewhat insulated from these pressures, is now also very much part of international benchmarking practice. However, differences across national research systems leave reliable international comparison difficult to achieve. An international review of different national systems draws attention to these difficulties and suggests some options for an evaluation framework that take into account some of the international systemic differences.

Introduction: global development in S&T

These days, when Governments invest in science and technology they have expectations that they will receive a return on those investments. There are, however, many differences in the nature of these expectations and these vary over time. They vary through different government administrations and in different countries. Returns can be expected to produce in one way or other, through public or private good, definable contributions to national socio-economic development. The latter is assumed to convert to public good through national economic development. In general, these expectations can be summarised under four broad headings:

- the expectation of generating *new knowledge*;
- the expectation of *solving* scientific, technical or social *problems*;
- the expectation of *enhancing scientific capacity* (infrastructure) ; and
- the expectation of contributing to an *economic platform for development*

Reflecting these expecations, governments around the world are demanding greater accountability an efficiency in the use of public monies devoted to science and technology. The demands to demonstrate outcomes and effective use of funds has extended to research institutes, universities and research councils.

Through the 1990s there was growing pressure on research funding agencies in most countries to report on the return to society of publicly funded national research investments and on the specific ‘results’ or *outcomes* of that investment. In several countries, this requirement has become enshrined in legislation or administrative regulation. In the United States, the 1993 Government Performance and Results Act (GPRA) requires strategic planning and performance reporting for all government activities, including research (Cozzens, 2000).

In 1993, Congress passed and the President signed into law the Government Performance and Results Act (GPRA, or "The Results Act"), P.L. 103-62. The broad intent of the legislation is to enhance the effectiveness, efficiency, and accountability of government programs by directing federal agencies to more singularly focus their management efforts on the results that are achieved - and away from such traditional concerns such as staffing and activity levels. Under GPRA, agencies must set goals, measure performance, and report on their accomplishments. And they must also ask and answer some basic questions: What is the institution's mission? What are the goals and how will them be achieved? How can performance be measured? How can information be used to make improvements?

As research investments have grown worldwide, the questions around it have become more insistent. What is the public getting for its money? Can it get more? These issues are driving the steadily growing demand around the world for performance information for publicly funded research. The questions arise not from distrust of research, but from its opposite: high levels of expectation that science is crucial to future prosperity. The steady increase in outcome assessment activity in all industrialized and most developing nations is a product of these expecations. In the U.S., while the fundamental research agencies have let program evaluation languish, the mission agencies have quietly reinforced and improved their assessment processes.

The nature of expectations of return also varies according to discipline. Investments in pure mathematics, for example, carry different expectations in the nature of return for those outside the field itself than do investments in biotechnology, electrical engineering or the social sciences. In many cases mission agencies are disciplinary focused, at least at the broad level. Health and medical research is typically of this order with specialist national health research agencies having responsibility for funding and evaluating national research efforts. In order to focus the discussion for this paper I have focused specifically on that area of research - health and medical research and development (HMRD). Similar issues and lesson can also be drawn for other fields of research.

The Case of Health and Medical Research

The health and medical research and development (HMRD) community has perhaps been more insulated than have other fields of science, particularly the social sciences. Until recently, governments appear to have accepted incontestably that HMRD is a public good that is carried out to the highest professional standards. There has been strong community support for HMRD, and last but not least the medical establishment and researchers are very effective lobbyists. Even so, HMRD is now operating in an environment of greater public accountability which is affecting all government-run or government-funded bodies.

Agencies such as the National Institutes of Health (NIH), which support fundamental research, now produce annual performance plans that assess the institutes' achievement of its performance targets. Most of NIH's performance goals are expressed objectively or quantitatively; other goals are assessed through an 'alternative form' such as descriptive criteria (Ordóñez- Matamoros, 2003).

Most evaluation expectations and processes seek to report on inputs, outputs and outcomes. Outcomes in this context refer to longer-term impact or effect expected or intended in a particular policy arena. Outputs are the immediate 'deliverables' — the goods and services — produced by the agency (Department of Finance and Administration, 2000).

As one moves from outputs to impacts, the results of research activity are generally broader in their effect, take longer to manifest themselves, are harder to quantify and are less readily traceable to particular research projects, funding programs or agencies. This is the so-called 'attribution factor'. While desired impacts may be seen, they have a plurality of causes, the individual contribution of which is not readily measurable. For example, while it may be possible to demonstrate statistically a decline in mortality or morbidity from a particular disease, it is in most circumstances impossible to attribute this with any confidence to a single cause, such as a major research program on the topic. There is an inherent difficulty too, as US researcher Paul David notes in respect of the US NIH, in requiring research agencies to consider 'outcomes beyond the span of control of the agencies which are being asked to define their outcomes' (Feller, 2002).

International performance indicators

In the case of Australian HMRD, the results of performance evaluation have usually been used to justify greater investment and targeting of that investment. As the primary funding body, the NHMRC has been developing an outcomes evaluation model for several years. International benchmarking has been part of this. In reviewing inputs, outputs and outcomes in Australia the NHMRC has engaged in two separate but related tasks.

1. *Benchmarking HMRD performance* The first has been to compare the performance of Australia's HMRD with that of the HMRD sector in other selected countries. However, this seemingly simple question raises many issues about the appropriate comparisons to attempt and the availability of data to support these comparisons.

2. *Benchmarking the use of performance indicators* The second task has been directed at learning from international experience with HMRD performance indicators, particularly with regard to those used by medical research funding organisations with similar functions to NHMRC.

Some interesting lessons can be drawn form this international benchmarking approach. They concern the feasibility of international benchmarking and how international performance evaluation might be improved in other fields.

Benchmarking of HMRD

‘Benchmarking’ can encompass many types of comparison of performance (Bureau of Industry Economics, 1996), but frequently implies a quantitative comparison. Benchmarking may be carried out for the purpose of research evaluation, but has wider application in policy analysis — e.g. for comparing inputs or investments. Research evaluation is concerned, literally, with assessing the value of the work of a particular researcher or research group, department, institution, agency or program. The focus is on the *results*, efficiency and effectiveness of the activity rather than simply on its cost or use of resources. The aim of research evaluation is to demonstrate *to a particular audience* the appropriate performance of the activity in question — Irwin Feller uses the term ‘multiple publics’ in this context (Feller, 2002). This implies a requirement to tailor the performance measure to that audience, and to use the sorts of evidence that the audience finds convincing (Garrett-Jones, 2000). A research council like NHMRC occupies a crucial ‘middle level’ (between government central agencies like DoFA and the researchers) in assessing the performance of research. In developing and using performance indicators, it has to talk to both the sponsor and the performer of research in language that they understand, using measures of performance that they find convincing. These measures may be purely qualitative, and thus not amenable to ‘benchmarking’. In summary, research evaluation and benchmarking are concepts that intersect but which are not identical.

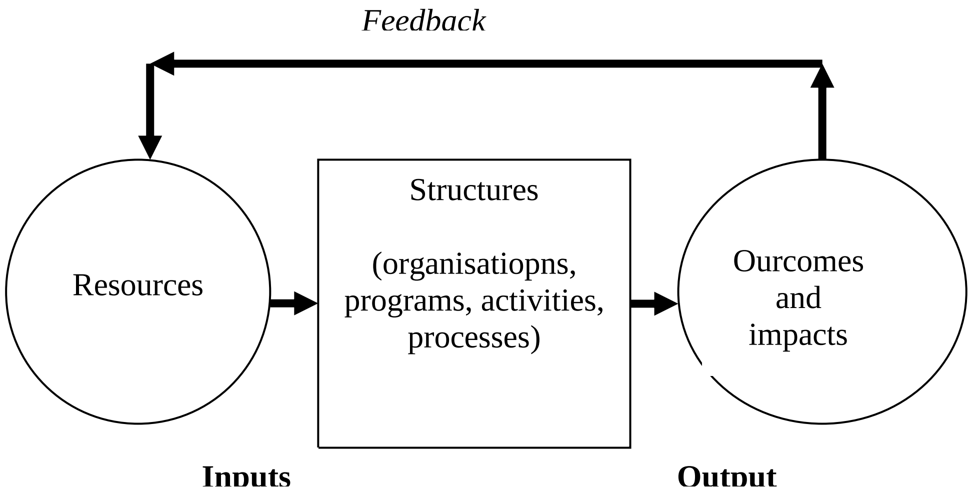
How do we apply them to evaluating HMRD? First, we can look at the components the HMRD system and consider how elements of the system might be used to ‘benchmark’ performance between countries.

Figure 1 presents a stylised model of a research system that might be applied to HMRD. Resources (recurrent funding, capital stock, personnel, existing knowledge, etc.) are the *inputs* to institutions or programs (research institutes, hospitals, universities, research councils) which ‘convert’ them into research *outputs* (publications, research trained people, patents, etc). The outputs produce beneficial health or other socio-economic outcomes or impacts. The challenge for research evaluation is to be able to tie the outcomes/impacts sufficiently to the structures and institutions to be able to identify the better performing institutions, programs or activities. These findings may influence the allocation of resources through some kind of feedback mechanism. In Australia, the federal education department’s funding formula for universities which takes account of various research ‘outputs’ is one example.

One form of research benchmarking is the qualitative international comparison often made by highlevel peer review panels. Here, the performance of a research discipline or institution may be ranked subjectively on the basis of its perceived international standing in adequacy of resources, effectiveness of structures and quality of outputs and outcomes. In the USA, NIH and the Food and Drug Administration have been involved in international comparisons by expert panels of the status of emerging areas such as tissue engineering. In Australia, one can point to the Australian Research Council’s reviews of grant outcomes — for example, in molecular biology — and the review of the Australian National University’s Institute of Advanced Studies (which includes the John Curtin School of Medical Research) as instances of qualitative benchmarking (ARC, 1994; 1996).

Figure 1: A Stylised model of a research system for evaluation purposes

(Following Arnold and Guy, 1998)



Defining the HMRD system

The first issue is exactly how to define the scope of HMRD, as represented by the central box in Figure 1. In practice it is quite difficult to come up with an acceptable definition for the 'structures' box in Figure 1. The OECD uses the term 'non-market R&D' to cover the public sector and non-profit private (PNP) foundations, but to exclude commercial HMRD performers. Alternatively, comparisons may be based upon government budget data (again as the OECD does) or funding, or on expenditures in or by particular agencies.

An alternative approach is to try and compare specific sub-sectors or agencies — for example, health research councils in different countries, or various schemes for supporting collaborative research.

In summary, the first problem in benchmarking is that structures, institutions and organisations are specific to the sector or country in question. Their performance can be measured in many cases, but these measurements cannot be directly compared with the situation in other sectors or countries.

Classifying the purpose of HMRD

One way to overcome the constraint imposed by specific local structures and institutions is to adopt a functional classification for R&D carried out in all sectors (i.e. businesses, public institutes, universities, hospitals). This would allow, as an example, a comparison of effort on reproductive medicine between countries. The federal statistics agency in Australia classifies all R&D by socio-economic objective (SEO). The SEO subdivision for 'health' covers 'R&D directed towards human health, including the understanding and treatment of clinical diseases and conditions and the provision of public health and associated support services' (Australian Bureau of Statistics, 1998). The subdivision has three groups: clinical (organs, diseases and abnormal conditions) public health, and health and support services. Comprising these groups are 49 classes, some of which refer to specific diseases or medical specialisms (e.g. endocrine organs and diseases (incl. diabetes); health related to ageing; diagnostic methods). R&D related to human pharmaceutical products and medical instrumentation is covered separately within the manufacturing subdivision of the classification.

Australia is fortunate in the degree of detail provided in its functional (SEO) classification of R&D. Regrettably, such comprehensive SEO classifications are not widely used.

The second problem, then, is the lack of an internationally accepted, detailed functional classification of R&D. This seriously limits the scope for international benchmarking of HMRD although, for the academic sector and for research publications, use of classification by field of research is an alternative approach.

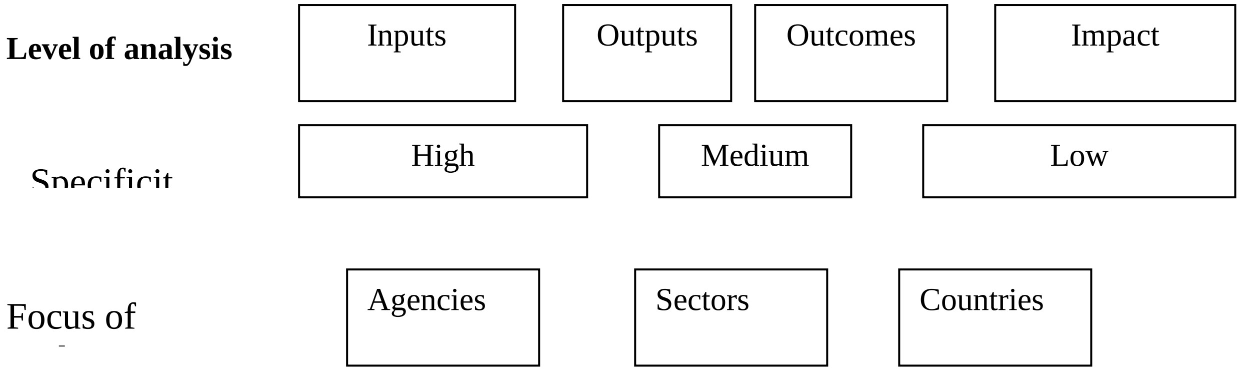
Measuring inputs, outputs and outcomes

Inputs and outputs can be counted more readily than outcomes and impacts. Research *inputs* (personnel and funding) are usually quantifiable. For most inputs the indicators are internationally standardised; for example, through adherence to the OECD's 'Frascati family' of guidelines and manuals. International standards or practices also exist for the measurement of many research *outputs* (publications, qualified people, patents), through firms like the Institute for Scientific Information (ISI) or agencies like national or regional patent offices.

Outcomes and impacts, on the other hand, often cannot be sensibly quantified and their assessment rests on qualitative criteria. A notable exception is the widespread use of impact measures of HMRD publications through analysis of the frequency of citations to papers in the international literature. International comparison is facilitated by ISI's databases and common classifications of research fields. It can be argued that bibliometrics provides evidence of the scientific impact of research rather than of its contribution to health outcomes. Given the time lag and the often indirect contribution of HMRD, attributing particular health outcomes to particular research programs is likely to be infeasible in most cases.

The third problem, then, is the difficulty of comparing research *outcomes* between agencies, sectors or countries. It must be recognized that the specificity of evaluation varies as the focus of analysis moves from agencies to countries and from inputs to impact (see Figure 2). Benchmarking of HMRD is therefore fraught with more uncertainty than are comparisons of 'hard' health infrastructure (e.g. hospital beds per 1,000 population in different countries). Because we can measure them readily, it is more feasible to benchmark R&D inputs and outputs than it is to benchmark outcomes. But there remain substantial difficulties even in comparing the resources available to health research in different countries.

Figure 2: Specificity of levels and foci of analysis



International approaches

Faced with these difficulties, what are other countries doing in HMRD evaluation methodology and practice? A selection of HMRD funding and research agencies from various countries (Table 1) is used to illustrate some differences and commonalities in the current use and development of performance indicators. Agencies faced with reporting on research (and research training) outcomes have adopted closely similar hierarchical approaches that

- identify the top level objectives that are to be achieved;
- show how the objectives of the research groups, centres, projects or institutions relate to the top level objectives; and
- specify available or potential indicators that can measure performance against both top level and research group objectives.

About half of the agencies reviewed in Table 1 either publish such performance indicators, or are developing or planning to develop them. Our review of the international data revealed several common features in the evaluation systems used by HMRD agencies.

- Because of the relatively recent history of HMRD performance evaluation there is less international standardisation than in other sectors of research. However, the study found no fundamental difference between systems that aim to measure the performance of HMRD and those for assessing research aimed at other socio-economic objectives.
- HMRD evaluation systems commonly take a hierarchical approach, linking top level (government or agency) objectives to outcomes that are valued at the research level, and to specific indicators that can inform the assessment of these objectives and outcomes.
- The better performance measures systems integrate *qualitative* and *quantitative* measures of performance and *internal* and *external* assessment. A framework that relies primarily on external qualitative review is expensive and difficult to implement and the results are likely to be incomparable with those obtained by other agencies. Systems that rely on readily available, internally generated statistical data are easier to define and implement, but risk overlooking the assessment of quality and relevance that can be provided by external review.
- The long-term social impacts and health impacts of HMRD are commonly assessed through *qualitative* studies involving the potential beneficiaries rather than through standardised statistical indicators.

Table 1: Uses of performance indicators by international HMRD agencies

Country	Health and medical R&D agency	Status of performance indicators
Canada	Canadian Institutes of Health Research (CIHR), formerly the Medical Research Council (MRC) of Canada	Reasonable information available. Performance indicators being further developed
France	Institut national de la sante et de la recherché medicale (INSERM)	Limited information available on performance
Germany	Bundesministerium für Bildung und Forschung (BMBF) – the Federal Ministry of Education and Research	Good information on the structure of the health and medical R&D system but availability of performance indicators limited
The Netherlands	Netherlands Organisation for Health R&D (Zon-Mw), incorporating the former Zorg-Onderzoek Nederland (ZON) – Health R&D Council and Medische Wetenschappen van NWO (MW-NOW) – Medical Sciences-Netherlands Organisation for Scientific Research	Does not collect information on publications, patents or commercialisation activities
New Zealand	Health Research Council (HRC)	Annual 'Progress and Achievements Report' gives qualitative and quantitative performance indicators. Funding a bibliometrics study of publications arising from HRC grants
Singapore	National Medical Research Council (NMRC), Ministry of Health	Little relevant information found. Current output reported does not allow for useful comparisons with Australia
South Africa	Medical Research Council (MRC)	Designing a performance indicators system
Switzerland	Swiss National Science Foundation (SNSF)	Financial (research expenditure information only). Annual report does not report on performance indicators and other publications do not appear to carry such material
United Kingdom	Medical Research Council (MRC)	Readily accessible information on performance. Performance indicators being further developed
United States of America	National Institutes of Health (NIH)	Several sources of information including annual performance plans and reports. Performance indicators being further developed.

Benchmarking HMRD

While many agencies have or are developing sophisticated performance measures, the issues outlined earlier mean that there is relatively sparse data. Some examples of feasible comparisons are given below. Each benchmark comprises a set of defined indicators (based on inputs, outputs and less commonly outcomes). As noted, the available comparisons are predominantly of research inputs and outputs, rather than outcomes. Comparisons are found at one or more of the following 'domains' or levels of evaluation: international, national, program or agency benchmarks.

Intensity of HMRD spending

One option is to compare the intensity of national HMRD as a proportion of the national expenditure on health care. Such data can be based on government budget expenditure for the main public HMRD agencies. As noted earlier, 'budget' data are incomplete as a national picture, but as health care expenditure is also budget-driven, the comparison may be appropriate. Table 2 offers a snapshot of how such comparative data can be arranged.

Functional priorities in HMRD

Despite the problems of incompatible functional classifications of HMRD it is possible to generate a general breakdown of research investment by health objective for the major research funding agencies in selected comparator countries. This can provide, albeit incomplete,an indication of the research priorities adopted by HMRD agencies across a range of countries. Agencies in several countries spend similar proportions of their funding on AIDS research, and the same pattern was true for the categories of mental health and infection/immunity. Table 3 gives an idea of the value of benchmarking using a detailed classification of the objective of HMRD. Another way of examining the relative national priorities in HMRD, at least for the more fundamental research, is through a country's share of world publications in relevant fields based on journal and article classifications. Table 4 shows how this comparison can proceed.

Table 2: HMRD 'budget' funding as a proportion of national health care expenditure

Country	HMRD as % of health care expenditure	Institutional HMRD funding per capita (US$)
USA	1.50	$64.06
Singapore	1.24	$8.14
New Zealand	0.54	$6.40
UK	0.53	$8.92
Canada	0.41	$8.02
France	0.39	$8.82
Australia	**0.38**	**$6.89**
Netherlands	0.35	$7.53
Germany (Est. b)	0.31	$8.43
South Africa	0.31	$0.70
Switzerland	0.30	$11.33
Germany (Est. a)	0.20	$5.50

Source: Turpin, Wixted and Garrett-Jones, 2003

Table 3: Agency funding by health area (percentage of identified R&D expenditure

Health objective	Australia 2002	Germany 2002	NZ 2001	Canada 2001	UK 2001–02	USA 2002	Singapore 2000
Cancer, cancer prevention and related disorders	**14.9**	5.4				21.0	
Cardiovascular health and diseases	**11.0**	3.7				8.7	
Endocrine diseases and diabetes	**4.2**						
Injury	**3.2**		3.6				
Mental health and neurosciences	**16.0**	13.2	15.3		16.7	25.3	
Respiratory diseases	**4.9**						
Bone, joint and muscle diseases	**4.0**						
Human genetics and inherited disorders	**3.8**	14.2	17.1				
Infection and immunity	**16.4**	13.4			16.7		
Liver, kidney and gastro-intestinal health and diseases	**4.9**						
Other health issues, diseases and conditions	**10.8**						
Reproductive health	**4.3**						
Social and environmental health issues	**1.6**	1.2					
Population groups and health		0.8	6.0		14.0		
Health sector/ system management		3.9	6.2				
AIDS research				10.7		10.7	10.7

Source: Turpin, Wixted and Garrett-Jones, 2003

HMRD priorities in relation to burden of disease

The federal and state governments in Australia have nominated seven national health priority areas based upon considerations of burden of disease and potential for improved health outcomes. Research priorities within NHMRC have reflected national health priorities and, more recently, the national research priorities (NRPs). 'Promoting and maintaining good health' is one of four NRPs announced in December 2002. Four specific goals — infant and child health, ageing, preventive healthcare, and the social and economic aspects of health — are identified as contributing to this priority. Research into biotechnology and genomics is included under other priority areas.

In response to the NRP initiative, NHMRC established strategic research networks (SRNs) in each of three health-related areas: 'Healthy Start to Life', Ageing Well', 'Ageing Productively' and 'Preventive Healthcare'. The Council has borrowed the concept of 'Consensus Conferences' from the US NIH with a view to developing further SRNs (NHMRC, 2003b). The Council has also identified priority areas, most notably Aboriginal and Torres Strait Islander health.

Unfortunately, the ability to benchmark these investments is severely limited by the lack of a common international functional or 'disease-based' classification for HMRD expenditures.

Table 4:Ranked Comparison of published papers in HMRD fields (papers/million pop. 1999)

Clinical medicine		Biomedical R&D		Health R&D		All fields of research	
Sweden	383.5	Switzerland	162.5	Australia	12.6	Switzerland	979
Switzerland	343.6	Sweden	154.2	UK	11.3	Sweden	940
Finland	328.1	Denmark	139.0	Finland	10.1	Finland	779
Denmark	287.3	Finland	113.0	Canada	9.7	Denmark	776
Netherlands	253.0	USA	101.8	Sweden	9.4	UK	667
UK	226.9	Netherlands	101.0	USA	9.0	**Australia**	**661**
Norway	202.0	Canada	100.7	Netherlands	7.3	Netherlands	660
Austria	197.8	UK	96.1	New Zealand	6.9	Canada	646
Australia	**197.1**	**Australia**	**89.3**	Norway	5.8	New Zealand	623
USA	192.9	Belgium	76.2	Denmark	3.9	USA	599
Canada	192.4	Norway	72.8	Switzerland	2.9	Norway	582
New Zealand	162.7	France	70.1	Belgium	1.4	Belgium	479
Belgium	160.5	Germany	67.7	Germany	0.9	France	455
Germany	134.5	Austria	62.4	Austria	0.9	Germany	454
France	126.0	New Zealand	51.7	France	0.5	Austria	442

Sources: Turpin, Wixted and Garrett-Jones, 2000; after National Science Foundation, 2002; population data from OECD, 2002.

HMRD outputs and impacts

Ideally, the 'priorities' data in Table 3 (which measures funding inputs) should be compared with the performance of each field as measured by outputs, such as publications or patents, or impacts such as citations to HMRD papers. Table 5 presents a comparative view of national contribution to health patents. Further work can be carried out by mapping the citations recorded for such patents.

HMRD outputs benchmarked against other fields of R&D

There are few cases where HMRD activities can be compared directly with similar international activities in other fields of research. However, it is possible to benchmark citation rates between different fields of research. Caution needs to be taken with respect to comparative levels of funding and 'expected' citation rates based on international experiences.

Table 7: Health patents by OECD country of inventor, all years 1988-1995

Country	No. of health patents (weighted)	% of world health patents
USA	22,409.4	55.8%
EU	10,810.1	26.9%
Japan	4,292.2	10.7%
Germany	3,030.5	7.5%
France	2,301.1	5.7%
UK	2,225.5	5.5%
Sweden	1,024.3	2.6%
Switzerland	759.5	1.9%
Italy	716.6	1.8%
Canada	684.2	1.7%
Netherlands	448.2	1.1%
Denmark	311.6	0.8%
Belgium	243.0	0.6%
Australia	224.3	0.6%
Austria	212.5	0.5%
Spain	124.2	0.3%
Finland	121.6	0.3%
Norway	97.5	0.2%
Korea	80.3	0.2%
Hungary	58.3	0.2%
Ireland	57.0	0.1%
New Zealand	31.4	0.1%
OECD	39,558.6	98.4%

Source: Lichtenberg and Virabhak, 2002

Conclusions and options for further development

These observations reveal something of a paradox about internationally benchmarking the performance of HMRD. On the one hand, even though performance evaluation of HMRD is a relatively new area, many international HMRD agencies have developed formal frameworks for reporting research 'results', or are doing so. On the other hand, there is rather limited scope for benchmarking of performance between countries. The prevalence of quantitative reporting on the basis of local organisational structures and categories makes even comparison of HMRD inputs difficult.

It is important to stress that, while there is ofcourse potential for science agencies to learn from the experience of other countries and other fields of research of performance evaluation — such as recent critiques of the influence of the US GPRA (Feller, 2002) — the lack of comparability is not because we are dragging our heels. What, then, is required to improve the evaluation and international benchmarking. Some possible avenues for further development can be noted.

.

A more structured approach

Without constructing a Byzantine evaluation framework, first consideration must be given to the *objective* of the benchmarking and the appropriate indicators to include. This reiterates an earlier comment on tailoring the evidence to the audience, in order to satisfy the 'multiple publics' for research evaluation. Both statistical indicators and measures of quality have a role in this process. In reviewing international practice the study observed four levels or 'domains' of benchmarking, as follows.

A. *International benchmarking* — data that attempt to reflect the performance at the national level (e.g. research funding, publications, citation analyses and commercialisation indicators, such as patents). Because of government reporting requirements, the emphasis to date has been on publicly funded R&D. There is a need to incorporate the HMRD activities of business and the nongovernment sector to give a full a national picture.

B. *National benchmarking* — benchmarking different national agencies (e.g. NHMRC and the CRC program) serves to assess their contribution toward national priorities. Here, some commonality of performance indicators between agencies would assist benchmarking, provided they were consistent with agencies' goals.
C. *Agency/institute benchmarking* — performance measures in this category would include areas such as the administrative cost and efficiency of administering programs, as well as some aggregate of the program benchmarks described below.
D. *Program benchmarking* — NIH's program level benchmarks using descriptive performance assessments and independent expert reviews provides a good example of this approach. Agencies such as NHMRC might record, on a regular basis, systematic information on qualitative outcomes. For example, grant recipients could be asked, on an annual basis, to identify and describe: (i) their most significant research breakthrough and (ii) their most significant health/medical outcome.

Each of these 'domains' contributes to an appropriately balanced portfolio of performance measures for a HMRD agency.

Common classifications?

As noted above one significant impediment to benchmarking is a lack of standardisation in the classifications used internationally for HMRD. This makes it hard to compare rigorously even *inputs* to HMRD between countries. Various different classifications are used for university and business research, and for inputs and outputs. Often, data are expressed in terms of a hybrid classification of organizational units, health specialisations and specific diseases or other health problems. The latter are more likely to be of interest in assessing the contribution of HMRD to improved health outcomes and reducing the quantified burden of particular diseases (Mathers *et al*, 1999). But, when it comes to the benchmarking the outcomes of HMRD, comparable international data are almost non-existent.

International benchmarking of HMRD performance would certainly be simpler and more robust if a common international and cross-sectoral (public/ academic/ private) detailed classification of HMRD objectives (SEO), like the Australian one, were to be implemented. This is, however, unlikely to come about quickly and, even if it did, would take us only part of the way towards comparing HMRD outcomes. The goal of these classifications is to specify the intended beneficiary of the research, and the categories used are not necessarily those which would be most helpful for tracking outcomes.

The more fundamental issue remains that of assessing the impact or outcomes of HMRD in terms which make sense to the researchers and stakeholders involved, and expressing these in ways which bear international comparison. Performance evaluation will always require a mix of statistical indicators and more qualitative, descriptive information on and expert assessment of research accomplishments. It is very hard to 'benchmark' the latter, although it can sometimes be put in semiquantitative terms (e.g. proportion of research objectives met or not met). Any assessment of the outcomes of HMRD is a two-stage process. It involves identifying the desired health outcomes, and then endeavouring to assess the contribution of HMRD to achieving those outcomes. In other words, benchmarking of HMRD performance is informed by, and in turn informs, the development of *health outcomes* goals and indicators. It must involve both the health and medical research community (who are the best judges of research quality) and the users and practitioners (who are best qualified to assess the impact and application of the research findings).

Perhaps, when proposing international benchmarks for HMRD outcomes, a more targeted, collaborative approach is therefore required. Clearly, there are many specific health issues (areas like HIV/ AIDS or mental health) where improved outcomes are of vital concern to many countries. International cooperation to define these specific areas and to track the contribution of HMRD to improved outcomes is one way forward which is likely to be acceptable to many countries. Success in this approach might prompt a closer alignment of the more general R&D classification with categories used by the health outcomes evaluation community.

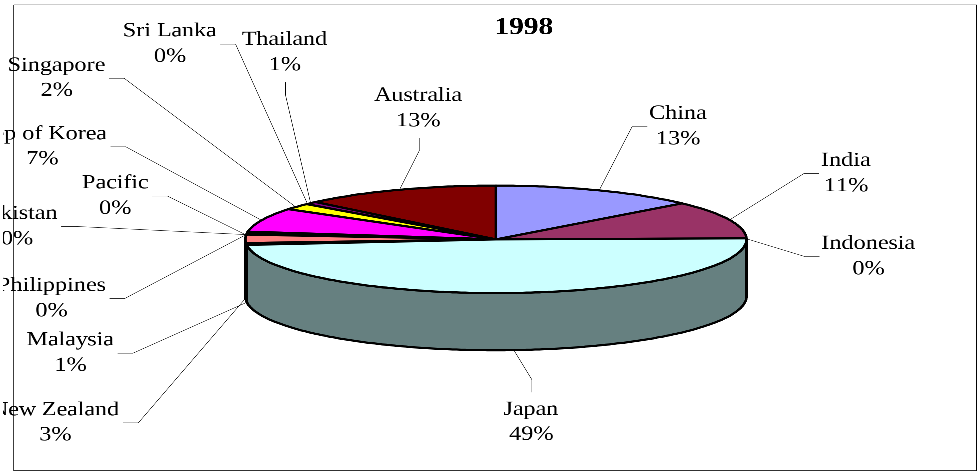
In all fields of research cooperation and coordination in the search for new knowledge and in its diffusion has become critical and all countries need to tap into global science as almost no single country has the resources to reach cutting edge level s in all relevant discipline and inter-disciplinary fields. A national sciences, technology and innovation system needs to be structured so as to participate as effectively as possible in global knowledge creation and diffusion networks. This is particularly important for economies making a 'transition' to a new economic base.

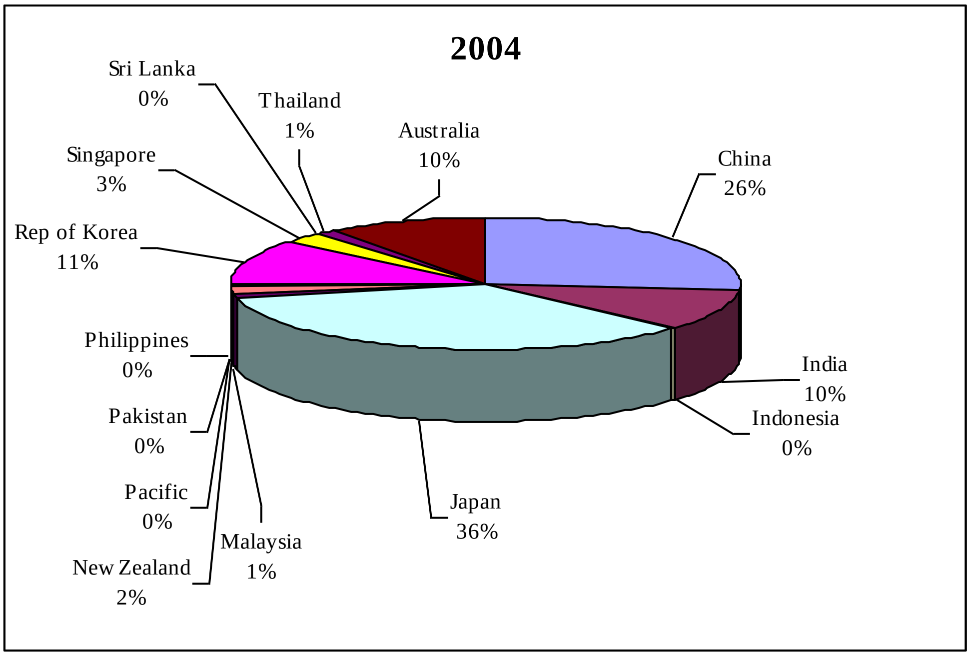
The 21st century now presents an industrial landscape dominated by strategic networks. These networks draw together skills and knowledge about science, technology, products, services and business systems into 'innovation webs'. Key actors in the networks include firms, research institutes, training institutes and other knowledge intensive institutions such as universities and hospitals. The innovation webs are constantly changing as the dominant players make decisions as to whether to expand or reduce access to the networks, whether to open entry to new members or to 'pull up the drawbridge'. But most significantly these innovation networks transcend national boundaries. In this context strategies for gaining and maintaining access to human resources for key technologies *in global demand* need to be collaborative rather than competitive.

The capacity of 'wizards' to deliver great outcomes and potential wealth can be convincing *only in the absence* of definable and reliable performance indicators and collaboration. International experience has shown that the capacity of science to deliver social and economic benefit can, in contrast, best be defined, understood and marshaled with the assistance of sound evaluation indicators and collaboration.

Additional Figures and Tables

ISI English language Journal Output for selected countries: 1998 and 2004





Example of Inputs, Structures, Outputs and Outcomes as identified for the Australian CRC Program, 2004)

Inputs	Structures	Outputs	Outcomes
Research staff Technical and support staff Research student enrolments Resources (staff, students, funding) by SEO and by subprogram Cost per staff member	Industry participation (no. offirms, funds) No. of students co-supervised by industry No. of participants in CRC (core/ supporting/ other) Level and growth in funds and contributions from participants Proportion of levered funds to CRC program funds No. of strategic international alliances	Contracted research and consulting (cases and income) Technology agreements, spin off companies Patents held and filed in Australia and overseas Technology transfer and professional training courses held (no. and income) Research student completions	Qualitative statement of achievements and outcomes Defined successful outcomes in research, teaching and commercialisation Return on investment case studies Qualitative account of most successful research outcomes No. of students taking up employment with industry

ISI English language Journal Output for selected countries: 1998 and 2004

Year	2004	2003	2002	2001	2000	1999	1998	Increase 1998-2004	% inc 2004 over 1998
China	53,235	46,900	38,469	33,206	29,004	23,398	18,833	34,402	182.7
India	20,867	20,803	18,525	17,501	15,983	17,104	16,037	4,830	30.1
Indonesia	456	489	395	486	422	381	342	114	33.3
Japan	70,865	78,046	71,207	70,215	69,773	70,435	68,585	2,280	3.3
Malaysia	1,257	1,171	934	922	814	869	798	459	57.5
Mongolia	60	56	41	39	45	38	34	26	76.5
New Zealand	4,046	4,135	3,793	3,842	3,847	3,798	3,793	253	6.7
Pacific	105	130	108	101	108	141	100	5	5.0
Pakistan	904	763	691	531	596	577	601	303	50.4
Philippines	424	440	410	315	351	344	311	113	36.3
Rep of Korea	22,031	20,529	16,642	15,519	13,200	11,894	10,458	11,573	110.7
Singapore	5,118	4,846	4,238	3,802	3,392	3,046	2,490	2,628	105.5
Sri Lanka	227	264	176	157	167	168	124	103	83.1
Thailand	2,049	2,048	1,591	1,331	1,185	1,043	935	1,114	119.1
Australia	20,985	16,573	21,601	18,968	18,486	18,898	18,302	2,683	14.7
All selected countries	***202,629***	***197,193***	***178,821***	***166,935***	***157,373***	***152,134***	***141,743***	***60,886***	***43.0***

References

B P J Alano and E S Almario (2000), *Tracking Country Resource Flows for Health Research and Development (R&D): A Comparative Report on Malaysia, the Philippines and Thailand with a Manual on Tracking Resource Flows for Health Research and Development* (Adarna House/Center for Economic Policy Research: Manila).

E Arnold and K Guy (1998), 'Technology diffusion programmes and the challenge for evaluation', *OECD Conference on Policy Evaluation in Innovation and Technology: Towards Best Practices, June 1997,* Paris (OECD).

Australian Bureau of Statistics (1998), *Australian Standard Research Classification 1998: A Set of Classifications for R&D Projects* (ABS: Canberra).

Australian Research Council (1994), *Reviews of Grants Outcomes No. 12: Molecular Biology 1987–1991* (AGPS, Canberra).

Australian Research Council (1996), *The Australian Council Review of the Institute of Advanced Studies* (AGPS, Canberra).

Bureau of Industry Economics (Australia) (1996), *Science System: International Benchmarking* (AGPS, Canberra).

L Butler (2003), *NHMRC-Supported Research: The Impact of Journal Publication Output 1996–2000* (NHMRC, Canberra).

L Butler and B Biglia (2001), *Analysing the Journal Output of NHMRC Research Grant Schemes* (NHMRC, Canberra).
.
L Butler, B Biglia and P Bourke (1998), *Australian Biomedical Research: Funding Acknowledgments and Performance* (NHMRC/ AusInfo.: Canberra).

Commonwealth Scientific and Industrial Research Organisation (2002), *Annual Report 2001-2002* (CSIRO: Canberra).

S E Cozzens (2000), 'Assessing federally-supported academic research in the United States', *Research Evaluation,* 9(1):pages 5–10.

Department of Finance and Administration (Australia) (2000), *The Outcomes and Outputs Framework Guidance Document* (DoFA: Canberra).

I Feller (2002), 'The good, the indifferent, and the less than attractive: II', *Research Evaluation,* 11(2): pages 95–99.

S Garrett-Jones (2000), 'International trends in evaluating university research outcomes: what lessons for Australia?', *Research Evaluation,* 8(2): pages 115–124.

S Garrett-Jones and T. Turpin (2002), *Measuring the Outcomes of the CRC Program: A Framework — Final Report* (Dept of Education, Science and Training: Canberra).

F Lichtenberg and S Virabhak (2002), 'Using patents data to map technical change in health-related areas', *OECD STI Working Papers*, 2002/16 (OECD: Paris).

C Mathers, T Vos and C Stevenson (1999), *The Burden of Disease and Injury in Australia* (Australian Institute of Health and Welfare: Canberra).

National Health and Medical Research Council (Australia) (2001), *Health and Medical Research Strategic Review: Implementation of the Government's Response — Final Report, October 2000* (NHMRC: Canberra).

National Health and Medical Research Council (Australia) (2002), *Investing in Australia's Health* (Focus Publishing: Edgecliff).

National Health and Medical Research Council (Australia) (2003a), *Performance Measurement Report 2000–2003: A Report on the Performance of the National Health and Medical Research Council 2000–2003* (NHMRC: Canberra).

National Health and Medical Research Council (Australia) (2003b), *The National Research Priorities: NHMRC Implementation Plan* (NHMRC: Canberra).

National Science Foundation (2002), *Science and Engineering Indicators 2002* (NSB: Virginia).

OECD (2002), *National Accounts of OECD Countries* (CD-ROM) (OECD: Paris).

G Ordóñez-Matamoros (2003), 'The US NIH performance assessment approach', in Turpin, Wixted and Garrett-Jones, pages 59–68. T Turpin, B Wixted and S Garrett-Jones (2003), *International Performance Indicators for Medical Research: Final Report to NHMRC* (Australian Expert Group in Industry Studies (AEGIS): University of Western Sydney).

P Wills (chair) (1998), *The Virtuous Cycle: Working Together for Health and Medical Research: Health and Medical Research Strategic Review* (the 'Wills Review') (Australian Government: Canberra).

A Young (2001), *Compiling Health Performer-Reported Health GERD: Measuring Expenditure on Health-Related R&D* (OECD: Paris).