RESEARCH ONLINE

University of Wollongong
Research Online

Faculty of Commerce - Papers (Archive)

Faculty of Business

2004

Some international benchmarks for evaluating Australian health and medical research

Samuel Garrett-Jones
University of Wollongong, sgarrett@uow.edu.au

Brian Wixted

Tim Turpin
University Of Western Sydney

Publication Details

Garrett-Jones, SE, Wixted, B and Turpin, T, Some international benchmarks for evaluating Australian health and medical research, Research Evaluation, 13(3), 2004, 155-166.

Research Online is the open access institutional repository for the University of Wollongong. For further information contact the UOW Library: research-pubs@uow.edu.au

RESEARCH ONLINE

Some international benchmarks for evaluating Australian health and medical research

Abstract
Recent experience in Australia has seen the requirement by the federal Department of Finance and Administration to conduct output pricing reviews of government agencies including research organisations. Health and medical research, while generally regarded as an important 'public good', is now pressed by the same demands as other research fields to account for public investments in terms of value of outcomes and value for investment. This paper reports on current trends towards international benchmarking of health and medical research performance. Comparative data from overseas show unique aspects of the Australian health and medical research funding system. The paper suggests possible future routes for carrying out health research evaluation in Australia.

Keywords
health, research, international, benchmarks, medical, australian, evaluating

Disciplines
Business | Social and Behavioral Sciences

Publication Details
Garrett-Jones, SE, Wixted, B and Turpin, T, Some international benchmarks for evaluating Australian health and medical research, Research Evaluation, 13(3), 2004, 155-166.

This journal article is available at Research Online: http://ro.uow.edu.au/commpapers/843

Research Evaluation, volume 13, number 3, December 2004, pages 155–166, Beech Tree Publishing, 10 Watford Close, Guildford, Surrey GU1 2EP, England

Comparative output reviews

Some international benchmarks for evaluating Australian health and medical research

Sam Garrett-Jones, Brian Wixted and Tim Turpin

Recent experience in Australia has seen the requirement by the federal Department of Finance and Administration to conduct output pricing reviews of government agencies including research organisations. Health and medical research, while generally regarded as an important 'public good', is now pressed by the same demands as other research fields to account for public investments in terms of value of outcomes and value for investment. This paper reports on current trends towards international benchmarking of health and medical research performance. Comparative data from overseas show unique aspects of the Australian health and medical research funding system. The paper suggests possible future routes for carrying out health research evaluation in Australia.

Dr Sam Garrett-Jones is an Associate Professor in the School of Management and Marketing, University of Wollongong, NSW 2522, Australia; email: sgarrett@uow.edu.au

Brian Wixted is Research Fellow and Professor Tim Turpin is Director at the Australian Expert Group in Industry Studies (AEGIS), University of Western Sydney.

This is an extended version of a paper presented at the conference on *Health Outcomes 2003: The Quest for Practice Improvement, 9th Annual National Conference*, Canberra, Australia, 20–21 August 2003.

GOVERNMENTS AROUND THE WORLD are demanding greater accountability and efficiency in the use of public monies for research. This pressure to demonstrate outcomes and effective use of funds has extended to research institutes and research councils. The health and medical research and development (HMRD) community has perhaps been more insulated than have, say, researchers in the social sciences. 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.

Demand for research performance evaluation

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). 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

 0958-2029/04/030155-12 US$08.00 © Beech Tree Publishing 2004

goals are expressed objectively or quantitatively; other goals are assessed through an 'alternative form' such as descriptive criteria (Ordóñez-Matamoros, 2003).

In Australia, federal government research agencies including the National Health and Medical Research Council (NHMRC) and CSIRO have since 1999 been subject to output pricing reviews by the Department of Finance and Administration (DoFA). These reviews examine the quantity, quality and price of outputs produced in an attempt to assess whether the government is getting good value for the taxpayers' funds. The process emphasises indicators of agency 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 NHMRC research program on the topic. There is an inherent difficulty too, as US researcher Paul David notes in respect of 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).

The Australian federal government framework therefore sensibly recognises that research agencies are valued as much for their standing *capacity* to deliver research expertise as for their specific R&D activities. In practice, the output pricing reviews of research agencies have been largely qualitative in nature (Commonwealth Scientific and Industrial Research Organisation, 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. The Council explains the rationale and objectives of the model thus:

> Scrutiny of the returns from the investment in research is one of the consequences of the introduction of an outcome-based accrual budgeting framework. The doubling of funding to the NHMRC, and the recommendations arising from the [Wills, 1998] Review, have encouraged the NHMRC to re-examine the way it performs its role as the peak funding body for Australian health and medical research … A key underlying theme of the Review calls for the NHMRC to develop an evaluation framework to quantify the outcomes of its investment in health and medical research … The Research Outcome Evaluation model … will form the basis of performance reporting for research funds managed by the NHMRC for the conduct of health and medical research. (NHMRC, 2001)

In the course of preparing its 2000–2003 Performance Report (NHMRC, 2003a), NHMRC therefore sought to present data and case studies on the outputs, outcomes and impacts of its activities. The Council asked the current authors to review and report upon performance indicators for HMRD that were in common use overseas (Turpin, Wixted and Garrett-Jones, 2003). The study had two main aims.

1. *Benchmarking HMRD performance* The first objective was 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 objective, therefore, was aimed 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. Here the review examined the state of development of international indicator systems for assessing HMRD, and considered how Australia's HMRD system in general and the NHMRC in particular compared with international practice in terms of the performance measurement frameworks being used.

The current paper briefly presents the main lessons learned from this study in the broader context of evaluation of HMRD. It explores, through examples, the feasibility of international benchmarking of HMRD and suggests how performance evaluation of HMRD in Australia could be further developed.

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, let us 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 high-level 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).

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 by particular agencies.

Table 1 shows the broad picture of HMRD expenditure and funding in Australia. From the table it is apparent that any analysis that considered only the Commonwealth government budget or the 'non-market' sector, for example, would only partially cover the HMRD sector. 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

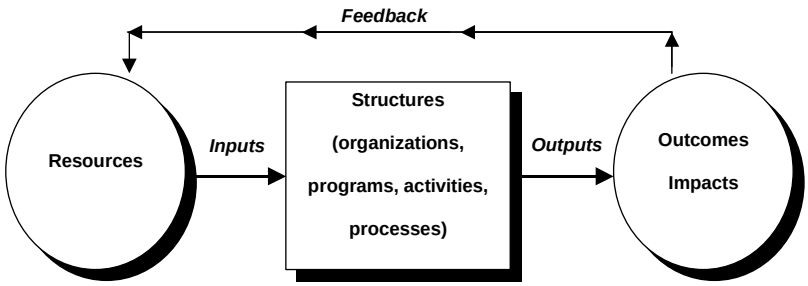

Figure 1. A stylised model of a research system for evaluation purposes, (after Arnold and Guy, 1998)

Table 1. Australian HMRD expenditure by sector of funding and sector of performance, 2000 (A$ mill.)

Sector providing funds	Sector carrying out R&D				
	Government research institutes	Universities	Other Australian (incl. PNP)	Business enterprises (pharmaceuticals only)	Total HMRD
Commonwealth govt research agencies' own funds	21.3				21.3
State govt research agencies' own funds	51.9				51.9
Universities' own funding (general university funds)		360.4			360.4
Business enterprises' own funds				198.6	198.6
Other Australian own funds			60.6		60.6
Commonwealth govt funding	27.7	130.5	44.8		203.1
Other Commonwealth govt funding for universities		39.6			39.6
Other schemes' funding for universities		6.8			6.8
State govt funding	18.9	27.8	20.6		67.4
Universities' funding	4.0		2.1		6.1
Business funding	27.6	37.9	14.2		79.7
Joint business/ govt funding	0.2		0.7		0.9
Other Australian (incl. PNP) funding	25.5	38.7	24.2		88.3
Overseas funding	5.6	26.1	13.6		45.3
Total HMRD	**182.7**	**667.7**	**180.8**	**198.6**	**1,229.7**

Source: Turpin, Wixted and Garrett-Jones, 2003

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 R&D 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 are covered separately within the manufacturing subdivision of the classification.

The performance of structures, institutions and organisations can be measured in many cases, but these measurements cannot be directly compared with the situation in other sectors or countries

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. As Alison Young (2001) comments:

> National specificities are particularly evident in the arrangements for non-market health R&D. National experts can mix and match institutional and functional data to reach a data set that they feel gives a reasonable picture of the level and structure of the R&D activities of their own National Health Science and Innovation Systems. *Only Australia took a purely functional approach, providing series based on socio-economic objective and on field of S&T.* Austria, Denmark, and to some extent Israel, also collected some project-level data in the non-market sector. Such detail was not available for Canada, France, the United Kingdom and the United States, which used semi-institutional approaches. R&D in the medical sciences was included for the university sector

> (incl. university hospitals). France and the United States also included some (United States) or all (France) other life sciences. [Emphasis added]

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. Analyses of the published outputs of Australian HMRD (Butler, 2003) use three separate 'field' classifications, for example.[1] This situation is likely to improve. Through the Health Research Systems project the World Health Organisation (WHO) has supported work aimed at describing and measuring health research systems in developing countries (Alano and Almario, 2000). The NHMRC is participating in the WHO project.

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.

Table 2 takes a real-life example and describes some of the indicators used to assess the performance of the Australian Cooperative Research Centres (CRCs). The program provides substantial funding for health and medical CRCs (Garrett-Jones and Turpin, 2002). These indicators are grouped under the headings of inputs, structures, outputs and outcomes. As can be seen, the outcomes indicators rely heavily on qualitative assessments or case studies, although separate bibliometric analyses have looked at the relative impact of publications from the health and medical CRCs (Butler, 2003). The outcomes performance measures used by NHMRC also rely on a mix of qualitative information (such as 'success stories') and bibliometrics (NHMRC, 2002; 2003a). The NHMRC has made extensive use of bibliometric analyses in benchmarking Australian biomedical research and to determine the scientific impact of the work that the council itself funds (Butler, 2003; Butler and Biglia, 2001; Butler *et al*, 1998). Bibliometric evidence was used very effectively in the Wills Review of health research in Australia (Wills, 1999).

The third problem, then, is the difficulty of comparing research *outcomes* between agencies, sectors or countries. 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.

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 3) is used to illustrate some differences and commonalities in the

Table 2. Some performance indicators used for health and medical cooperative research centres in Australia

Inputs	Structures	Outputs	Outcomes
Research staff Technical and support staff Research student enrolments Resources (staff, students, funding) by SEO and by sub-program Cost per staff member	Industry participation (no. of firms, 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

current use and development of performance indicators. It appears that international HMRD agencies are under the same pressures as NHMRC to report on the outcomes of the research and training activities they carry out or fund. Thus, many are implementing formal reporting of research 'results', usually through both statistical and qualitative measures of performance. Agencies faced with this task 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 3 either publish such performance indicators, or are developing or planning to develop them.

The 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.

Benchmarking Australian HMRD

While many agencies have or are developing sophisticated performance measures, the issues outlined earlier mean that there is relatively sparse data avail-

Table 3. Use of performance indicators by HMRD agencies internationally

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.

Table 4. HMRD 'budget' funding as a proportion of national health care expenditure and per head

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

available for international benchmarking of HMRD. 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 within the following 'domains' or levels of evaluation: international, national, program or agency benchmarks.

Intensity of HMRD spending

Table 4 compares the intensity of national HMRD as a proportion of the national expenditure on health care in Australia and 10 other countries. The data are 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. Some estimation is required, as in the alternative figures for Germany. HMRD spending per capita is given for comparison.

The analysis reveals strong similarities between the countries (except the USA), with most spending between 0.3% and 0.55% of their health care budget on HMRD. Australia falls within the mid range of the countries examined.

Functional priorities in HMRD

Despite the problems of incompatible functional classifications of HMRD, the study attempted a breakdown by health objective for the major research funding agencies in seven of the countries. Table 5 gives an (incomplete) indication of the research priorities adopted by HMRD agencies in these countries. Agencies in several countries spent a similar proportion of their funding on AIDS research, and the same pattern was true for the categories of mental health and infection/immunity. The table 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 a field (see Table 6).

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. These are: asthma, cancer control, cardiovascular health, diabetes mellitus, injury prevention and control, mental health, and arthritis and musculoskeletal conditions

Table 5. 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

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 a field

(Mathers *et al*, 1999). In 2003, over 60% of NHMRC funding was dedicated to research in these priority areas (NHMRC, 2003a).

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 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.

HMRD outputs and impacts

Ideally, the 'priorities' data in Table 5 (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 Australian HMRD papers. Table 7 shows Australia's ranking in published scientific papers in three HMRD sub-fields (clinical medicine, biomedical research, and health R&D) as defined by the US NSF. The countries chosen for comparison were the principal European OECD countries with Canada, the USA and New Zealand. Data are expressed as number of publications per head of population. Australia ranks in the middle of the pack for both clinical medicine and biomedical research. However, for its population size Australia publishes more research in the sub-field of 'health'. The four Scandinavian countries are prominent towards the top of the list for both clinical and biomedical research.

Data on publications can also be used as a measure of research impact. Using journals listed by the ISI for the period 1996–2000, Butler (2003) has investigated the citations attracted by published Australian biomedical papers against two benchmarks. The first compares the actual rate of citation to Australian

Table 6. Australian and world citation rates for biomedical papers, 1996–2000

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)

Notes: a. Fields omitted because of low publication counts are: clinical chemistry; dentistry; nursing; human movement and sports science; and other medical and health sciences.
b. This measure is equivalent to the relative citation impact (RCI) indicator described by Butler *et al*, 1998.
c. ISI captures only about 25% of Australian articles in *Public Health* and health services. In other fields, ISI captures more than 80% of articles.

Table 7. 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.

papers to the average citation rate for papers in the journals in which they are published (the 'expected' citation rate). The second, the relative citation impact, or RCI (Butler *et al*, 1998), makes comparison with the world average for citations to biomedical papers published in all ISI-listed biomedical journals. These two benchmarks may differ if, for example, the Australian papers are published in 'lower impact' journals, i.e. those with a lower average citation rate per paper, than is normal for the field.

Table 7 shows the results of this comparison for selected main fields of HMRD. Butler (2003) demonstrates that, for most sub-fields in Australia, the actual citation rate per paper approaches or exceeds the 'expected' citation rate. The sub-fields of general medicine, immunology and clinical sciences show the highest citation impact. Using the RCI, the fields showing an impact greater than the world average are immunology, clinical sciences, biotechnology and microbiology. Sub-fields like neuroscience and medical physiology return a relatively weaker performance against both indicators.

Overall, Butler's data (Table 6) suggest that the average citation rate of Australian biomedical papers is on a par with the international benchmarks, but there is a slight tendency for papers to appear in journals with a citation 'impact' that is lower than the world average. This suggests that fundamental HMRD is performing notably well by international benchmarks, but perhaps that Australia's HMRD system overall is skewed towards applied research (which tends to be less highly cited), at least by comparison with the publications emanating from the major European and North American research centres which largely determine the 'world' average.

Perhaps Australia's HMRD system overall is skewed towards applied research, at least by comparison with the publications emanating from the major European and North American research centres

While the data on publications and citations reveal that Australian HMRD has been performing strongly, the same does not appear true in patenting. A recent study of patenting in health-related fields (Lichtenberg and Virabhak, 2002) investigated patenting in the field 'Medical or Veterinary Science; Hygiene' — meaning pharmaceuticals. The authors used the OECD 'Triadic Patent Families' database, which consolidates raw patent data from the European, Japanese and US Patent Offices. Among the benchmark OECD countries only New Zealand ranked lower than Australia in terms of its share of world health patents (Table 8). Furthermore, time series data show that Australia's performance in patenting, as a proportion of world activity, is declining (Lichtenberg and Virabhak, 2002).

HMRD outputs benchmarked

One of the concerns that prompted the current study was whether research funded by NHMRC was achieving appropriate outputs and outcomes by comparison with R&D funded by other research councils. Butler (2003) clearly demonstrates that research supported by the NHMRC attracted citation rates that were greater than the Australian average and, in all but two sub-fields, were also above the world average (i.e. RCI > 1.0). The rate of both actual and expected citations was substantially higher for the more fundamental research supported by the NHMRC or carried out in specialist research institutes and cooperative research centres than for the biomedical publications from other research groups in government laboratories, hospitals or universities. As there are few cases where HMRD activities can be compared directly with similar international activities in other fields of research, Table 9 examines the

Table 8: 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

contribution of HMRD within the Australian Cooperative Research Centres Program. Health and medical research accounts for around 11% of the researchers within the CRC program. However, medical CRCs contribute a higher proportion of CRC outputs in several areas, notably in the numbers of patents and the value of technology agreements.

In summary, the range of 'benchmarks' of Australia's HMRD performance is limited, and several involve 'best estimates' of other countries' performance. One should be rather cautious about interpretation, especially as no measures of the quality or international recognition of Australian HMRD (beyond those of citation impact) are presented.

Table 9. Outputs from medical Cooperative Research Centres in Australia

CRC program performance measure	Contribution of medical CRCs to the program total (1991–92 to 2000–01 except where indicated)
Researchers (full-time equivalent)	10.6%
Papers accepted for publication	11.0%
Journal articles published, 2000–01	12.3%
Patents maintained in Australia	38.0%
Patents filed in Australia	31.9%
Patents maintained overseas, 2000–01	21.4%
Patents filed overseas, 2000–01	25.0%
No. of R&D contracts and consultancies	3.7%
Value of R&D contracts and consultancies (at 2001 prices)	12.6%
No. of technology agreements	11.8%
Value of technology agreements	27.9%

Source: Garrett-Jones and Turpin, 2002

However, on the data available, Australia's HMRD community seems to perform well, typically being ranked with the medium-sized European countries, Canada and New Zealand. There is little to separate these countries in terms of public resources for HMRD and they have rather similar research priorities. Australia ranks close to the UK and USA in published papers (per capita) in clinical medicine and biomedical R&D and ahead of both in health R&D. Australia also performs very well in terms of the impact of published HMRD. Where Australia appears less competitive is in the area of potential commercial outcomes of research. The number of Australian health patents is around half that of the Netherlands, for example — a country with a similar population — and our output appears to be declining. This may reflect the performance of the business sector, which is not well captured in the benchmarks presented.

Conclusions and future work

The current study revealed something of a paradox about benchmarking the performance of Australian 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 of course potential for Australia 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. Through NHMRC, Australia is keeping abreast of international best practice in HMRD performance evaluation. Further, with the comprehensive Australian Standard Research Classifications (ASRC) Australian statisticians are leaders in measuring detailed R&D objectives which help greatly in benchmarking R&D outcomes across the public and business sectors.

What, then, is required to improve the evaluation and international benchmarking of Australian HMRD? In conclusion, we put forward some proposals, pose some questions, and suggest possible avenues for future work.

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 our 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.

- *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 non-government sector to give a full national picture.
- *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.
- *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.
- *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: (a) their most significant research breakthrough and (b) 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 we have noted, 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 organisational units, health specialisms 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 semi-quantitative 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 classifications with categories used by the health outcomes evaluation community.

In conclusion, we would argue that further effort in assessing the outcomes and impacts of HMRD in Australia should be more closely connected with work that aims to develop health outcome indicators and measures of the health-related quality of life.

Acknowledgments

This research was funded under a NHMRC contract with AEGIS, University of Western Sydney. We thank Linda Butler, Susan Cozzens, Gonzalo Ordóñez-Matamoros and Alison Young for their assistance and advice.

Notes

1. These are the ISI subject category for journals, the ABS research field, courses and disciplines (RFCD) categories, and an internal NHMRC classification which more closely reflects clinical specialisms.

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).