1 Good public policy — why evidence and process matter

Effective policy development demands careful analysis of different options, drawing on available evidence. Good process is the key to ensuring that this happens, whether in developing new policies or evaluating existing programs. Evidence-based analysis and good process matter because getting policy right matters. Public policy measures can have pervasive effects on the wellbeing of the community.

The community deserves assurance that policies are designed and implemented to produce the outcomes it seeks in a cost-effective way. Reforms that raise the quality of spending can reduce taxation imposts and enable the objectives of regulation to be met more efficiently, contributing to growth in productivity, employment and income. In contrast, policies conceived without proper assessment carry risks of locking in productivity-sapping impacts and reducing the capacity to fund more worthwhile initiatives.

During the 1980s and 1990s, Australia gained an international reputation for policy processes and reforms which involved the assembly of evidence and transparent consideration of options. But there is scope to do better. In this chapter, the Commission draws on its experience in conducting inquiries and research studies over the years to identify ways of strengthening evidence-based policy development in the future.

Evidence and process — the two pillars

While good evidence and due process are fundamental to good public policy, they cannot assure it. Public policy is influenced by a variety of stakeholders, analysts and decision makers who will tend to interpret evidence through a particular 'lens' based on their own values, perceptions and interests. Governments must often make contentious policy decisions in a 'politically charged' arena, on behalf of a community with its own preconceptions but often with less access to relevant information.

These difficulties reinforce the need for good process — especially transparency to draw out and test the available evidence — and, ultimately, government leadership to promote and consolidate community support for reform. Drawing on the experiences of its member countries, the OECD (2010e, p. 9) observed recently that:

> … when advancing contentious reforms, experience suggests that successful leadership is often about winning consent rather than securing compliance. This makes effective communication, underpinned by solid research, all the more important.

Better use of evidence and sound policy processes will be crucial in advancing the Council of Australian Governments (COAG) reform agenda. This aims to raise workforce participation and productivity growth, and thus underpin rising living standards in the face of an ageing population. Under that agenda, all governments have endorsed ambitious goals and spending programs, much of which are in the 'data-challenged' field of human capital development (table 1.1).

Evidence-based analysis engenders support for reform

Balanced ex ante analysis can help make the case for well-conceived reform and careful ex post evaluations of policies can help secure better government decisions and consolidate support for them within the community.

Following a remit from COAG, an Industry Commission study (1995) projected that the National Competition Policy (NCP) could generate a net benefit equivalent to as much as 5.5 per cent of GDP if fully implemented. In a 1999 inquiry, the Commission similarly projected a boost in the level of GDP of 2.5 per cent from selected NCP reforms of relevance to regional Australia (PC 1999c, 1999d). Retrospective studies since then support the scale of these estimates. The Commission's 2005 review of the NCP identified that the realised productivity and price changes in key infrastructure sectors alone in the 1990s — to which the NCP had directly contributed — had increased Australia's GDP by 2.5 per cent, or $20 billion (PC 2005a, 2005h).

Similarly, Commission researchers modelled the potential benefits from allowing irrigators to trade water during drought (Peterson et al. (2004)). They found that allowing trade more than halved the impact of the reductions in water on the gross regional product of the southern Murray-Darling Basin by mitigating the losses in the activities most reliant on water. Subsequent analyses by the National Water Commission (2010), Frontier Economics (2007), Mallawaarachchi and Foster (2009) and Productivity Commission (2010a) confirmed that water trading in the Basin had enabled many irrigators to survive consecutive years of drought.

Table 1.1   **The COAG Reform Agenda: a snapshot**

Reform area	Objective/s	Some key priorities/initiatives
Health and Ageing	• Improve health outcomes and the sustainability of the health system	• Preventative health: reduce smoking, obesity and diabetes • Increase access to primary and community healthcare (GPs, dentists, mental health practitioners), hospitals and aged care • National registration and accreditation scheme for health professionals • More efficient pricing of public hospitals and health workforce reforms
Productivity	• Improve human capital outcomes through reform in areas of education, skills, early childhood development & teacher quality	• Lift year 12 retention rates, literacy and numeracy achievement • Improve teacher quality • National reporting on performance • National vocational education and training system
Climate change and water	• Ensure an effective national response to climate change, through an emissions trading scheme and nationally consistent set of climate change measures • Ensure sustainable water use across Australia	• Murray-Darling Basin Agreement • Reform national water markets • National strategy for energy efficiency • National renewable energy scheme
Infrastructure	• Improve infrastructure planning and investment • Remove blockages to productive investment	• Infrastructure Australia work program (national infrastructure audit, priority list) • Develop best practice guidelines for public-private partnerships
Business regulation and competition	• Reduce regulatory burdens on business • Delivering deregulation and competition priorities • Improve processes for regulation making & review	• 27 deregulation priorities including OHS, occupational licensing, food regulation, consumer policy and credit, environmental approvals process and payroll tax • 8 areas of competition reform including anti-dumping, parallel importation of books and national transport, infrastructure and energy reform.
Housing	• Improve housing supply and affordability • Halve the number of homeless people turned away from shelters within 5 yrs • Improve social and community housing	• Increase supply of land, access to social housing • Planning reform • National approach to homelessness
Indigenous reform	• Close the gap on indigenous disadvantage, particularly for: life expectancy, child mortality, literacy and numeracy	• Increase access to quality early childhood education, schooling, vocational education and health services • Reform the provision of social housing for Indigenous people • Improve community safety (target domestic violence, drug and alcohol abuse).

Sources: COAG Communiqués (20 December 2007, 26 March 2008, 3 July 2008, 2 October 2008, 29 November 2008), Rudd, Swan and Roxon 2010.

Quality ex post evaluation can also identify where and how policy may need refinement. A Commission review of the Job Network found that while the net employment outcomes were low, as with previous programs, Job Network achieved outcomes at lower cost. The Commission identified some adjustments to scheme design and administration that could improve the effectiveness of services for job seekers with little impact on funding costs (PC 2002e).

In other cases, evaluation can point to the need to terminate a policy or program. The NCP's evidence-based process put the onus of proof on those seeking retention of anti-competitive regulatory provisions to establish that such restrictions were in the public interest. As a consequence, hundreds of legislative restrictions that benefited particular interests, but imposed larger costs on the rest of the community, were removed (NCC 2005; PC 2005a).

The reforms that drove Australia's improved economic performance in the late 1980s and 1990s were aided by evidence of the economy-wide costs of protection and anti-competitive regulation, together with processes that used this information to build community support for reforms (box 1.1). For these earlier reforms, the OECD commented that other countries could learn from Australia's 'willingness to commission expert advice and to heed it, to try new solutions, and to patiently build constituencies that support further reforms' (OECD 2004).

Some policy areas, however, are resistant to quantification, with more qualitative analysis and judgment being required. In its review of executive pay for example, the Commission was hindered by a lack of consistent, long-running data on remuneration and a complex interaction of 'black letter' and 'soft' law that sought to achieve subtle behavioural responses in company boardrooms. It noted that 'with all the uncertainty, considerable judgment is called for, particularly in relation to the magnitude of identified problems and the relative downside risks in intervening versus doing nothing' (PC 2009g, p. 11).

Transparent public processes are important to ensure that necessary judgments by advisers and decision-makers can be adequately scrutinised and tested, particularly by those who will be affected. For more intractable policy dilemmas, there is a strong case for iterative policy reforms that draw on accumulating evidence along the way to help assess the direction of long term reforms. Confronted with uncertainties (and potentially large costs) in its review of road and rail infrastructure for example, the Commission recommended a phased approach to reform with each step preceded by examination of costs, benefits and distributional impacts (PC 2006h).

<div style="border: 1px solid black; padding: 10px;">

Box 1.1   Good evidence and process: 'reform-era' successes

Implementation of the transformative reform agenda of the 1980s and 1990s required skilful political leadership and community acceptance of change. Evidence-based processes assisted in a number of ways.

- They promoted awareness of the costs of existing policies and the benefits from reform
  - for example, the Commission's (and its predecessors') analyses of the costs borne by the mining and agricultural sectors as a consequence of manufacturing protection helped to galvanise those sectors as major political forces for tariff liberalisation (PC 2003c; Banks 2005, 2010).

- They gave governments the opportunity to gauge community reaction to reform options, reducing the prospect of unanticipated responses
  - for example, the Commission's inquiry into Private Health Insurance (IC 1997) examined the community rating system which prevented health funds discriminating on the basis of age. The system was initially perceived as both 'fair' and politically off limits. However, analysis showed that it actually was leading to inequities. With young people not contributing to the pool, premiums for remaining (generally older) members spiralled, resulting in further exits. The Commission's recommendation that people entering insurance late should pay higher premiums than those who enter early gained support and was adopted by the Government.

- They enabled governments to make the case more convincingly for policy changes and to resist pressures to introduce nationally costly measures
  - for example, the Commission's analyses of work practices, including in waterfront (PC 1998a, 1998b) and construction (PC 1999e), exposed cost padding and inefficient practices which provided independent support for calls for reform.

</div>

Building better evidence *through* good process

The Commission's inquiries and studies have revealed some key features of the policy process which help achieve good outcomes. Policy progress can be accelerated, and costly errors reduced, by:

1. clearly defining the problem to be addressed and establishing a conceptual framework to guide evidence gathering and interpretation

2. acquiring better data on 'baseline' situations and measuring the changed outcomes as new policies are implemented

3. consulting widely to ensure that all available evidence is incorporated, and providing early 'airing' of proposed policy options to test their viability

4. building quality evaluation into the implementation of policies and using evaluation to shape improvements

5. evaluating different policy approaches across states and territories

6. making effective use of scarce evaluation skills by drawing on academic expertise and sharing experience across jurisdictions.

Defining the problem clearly

Sometimes the reasons for public concern in a policy area and what the community expects to achieve through government intervention are unclear or imprecise . And community perceptions of the scale or urgency of a problem can sometimes run ahead of reality, or be heightened unduly by sectional interests. Failure to define a problem carefully can lead to unnecessary and inappropriately designed policy actions.

- Participants in the Commission's inquiry into paid parental leave variously identified nine different (and sometimes conflicting) objectives for a leave scheme. Analysis enabled those objectives which were unlikely to be affected by paid parental leave to be set aside, allowing a focus on those that were relevant to policy design (PC 2009h).

- The Commission was asked to assess Australia's consumer product safety system with the presumption that there were major problems. While areas for improvement were identified, the evidence suggested that rates of injury and death from faulty consumer products had been declining, and that remaining injuries were often attributable to consumer misuse or environmental factors, rather than to the products themselves (PC 2006d).

Sometimes policy objectives and policy measures can drift out of alignment. The Commission found that drought policy *objectives* — namely, preparing for drought, managing and coping during drought and recovering after drought — recognised drought to be a recurring feature of climate and sought therefore to promote self reliance. However, as they evolved over time, drought *programs* focused almost exclusively on crisis payments during drought, undermining such self-reliance (PC 2009c).

Better data

It is difficult to evaluate the impact of a policy without information about the state of play before it was introduced. But even good baseline data are of limited value if there is little information on how performance evolved subsequently. For a 1999

inquiry into gambling, the Commission found the data deficient to judge key policy dimensions of problem gambling, and therefore conducted its own survey to establish a baseline (PC 1999a). Its subsequent 2010 inquiry found that, while much data are now collected, they are not appropriately targeted to shed light on gambling policy issues and are marred by differences among jurisdictions (PC 2010b).

The challenge of accessing information that has been collected can compound the problems of deficient or missing data. Data are of limited value if few researchers have access. In a study commissioned by the Australian Government into public and private hospitals, the Commission encountered long delays in getting access to the data it needed (box 1.2).

Box 1.2   Accessing performance data on hospitals

For its study comparing the performance of public and private hospitals, the Commission found that datasets are limited by missing information and inconsistent collection methods — for instance, there is no robust data on the costs of public and private hospitals, and no nationally consistent data on hospital-acquired infections. Compounding this were significant delays in accessing data beyond what could reasonably be caused by a need to address privacy concerns.

The Commission encountered: requirements imposed by the Australian Bureau of Statistics (ABS) that restrict use of public data; barriers to accessing data held by the Australian Institute of Health and Welfare due to a need to obtain approval from state and territory governments for its release; and the need for private hospital operators to approve access to their data.

These data collections could be made more available to researchers to interrogate and to identify improvements in the efficiency and effectiveness of health care in a way that meets legitimate confidentiality restrictions. The community funds data collections in the public hospital sector — including collections compiled by the ABS — and, through public and private contributions to the cost of private hospital care, also contributes to the cost of data collections in the private sector. Accordingly, there is a strong case for maximising the benefits that the community achieves from data it has paid for.

Source: PC (2009f).

The Steering Committee for COAG's Review of Government Services — for which the Commission provides the Secretariat — has over the years found some government-funded bodies to be unresponsive to requests for information. In some cases, even new presentations of already published data have required Ministerial authorisation, and a single jurisdiction has been able to veto publication of the data.

Since 2002, the Steering Committee has been reporting on indicators of Indigenous disadvantage to inform governments about whether policies are achieving positive

outcomes. Although four reports have now been produced, there is still a lack of comparable trend data for about half of the 50 indicators (SCRGSP 2009, Banks 2009a).

Consultation and 'stress-testing' of proposals

Transparent consultations can help to identify the full impact of proposed changes in policy or regulations and can be an important source of evidence in their own right. Early consultation can establish a clearer understanding of the 'baseline' situation, the magnitude of problems, the extent of compliance costs and can ensure all relevant options are considered and possible unanticipated consequences uncovered. After initial consultations, draft reports, 'green papers' and exposure drafts can be used to 'stress test' preliminary policy proposals.

The Commission's latest review of regulatory burdens on business found that many in the finance and property industries considered the most significant regulatory failings to be: a lack of transparency and continuity in consultation processes, short consultation timeframes and a lack of credible evidence in current regulation-making (PC 2010c). When consultations are short and bound by confidentiality agreements, participants cannot be confident that their views have been properly weighted and that others' views have been tested appropriately.

During the Commission's inquiry into executive remuneration, employees, employers and unions raised concerns about a new policy initiative to change the taxation treatment of employee share schemes. Information received by the Government in response to the announcement, on which there had been little consultation, resulted in several revisions to the policy, with a final position emerging around two months later, during which time many such schemes were put in abeyance (PC 2009a).

The costs and time involved for a robust consultation process, while significant, are generally much smaller than the costs and uncertainty generated by ill-advised policy measures that subsequently need to be recast. Robust consultation in the early stages also avoids governments being placed on the back foot and needing to negotiate late changes with those affected. Such negotiation can lead to inferior outcomes.

Building quality evaluation into the policy

Better data will not of itself lead to improved policies; data needs to be put to use effectively. Evaluation sometimes seems an afterthought to the policy

implementation process and, on occasion, appears geared only to enable proclamation of success or failure rather than policy improvement. It is rare to find funding for and design of high-quality evaluation at early stages to guide policy implementation and refinement. The Commission's study into the not-for-profit sector recommended building into policy initiatives from the outset the mechanisms and funding for data collection and post-implementation reviews (PC 2010d).

One requirement for good policy-making is independent scrutiny of, and public reporting on, the performance of agencies in developing and implementing regulations. Most jurisdictions now require regulatory impact statements (RISs) for new or amended legislation. This encourages the adoption of benefit-cost frameworks. However, there is scope for improvement. In its recent annual reviews of regulatory burdens, the Commission has recommended that the Australian Government improve the transparency and accountability of its gatekeeping processes and that a 'consultation' RIS be incorporated into the regulation making process (PC 2009b, 2010c).

The RIS process is less well equipped to ensure effective post-implementation outcomes. The Commission has previously identified a need to improve the machinery for stronger monitoring of outcomes (PC 2005a).

Systematic learning across jurisdictions

New policies often cannot be implemented at the same time everywhere. The inevitable need for sequencing of project roll-out can be turned to advantage and yield important benefits if used to design and evaluate pilot projects. Differences in policy approaches across jurisdictions also enable policy learning from the natural experiments to which such variations give rise.

A positive example of 'competitive federalism' was noted in the Commission's review of the NCP. The rewards and sanctions of the NCP framework, in concert with public assessments, motivated governments to learn from different regulatory approaches in other jurisdictions to advance their own reform agendas (PC 2005a).

Outside of that formal incentive framework, learning through policy experimentation has been less common, particularly for social policies. Indigenous policy in particular, has suffered from this deficiency.

> One advantage of our Federation is that it has generated many different policy and program innovations. However, with some exceptions, Australia has squandered the opportunity to learn systematically from these diverse experiences in order to identify those that could make a difference if applied nationally. (Banks 2009a)

The Steering Committee for COAG's Review of Government Services has sought to redress this in its work by including case studies of programs that are, or appear to be, working. In its reporting on governments' service provision, the Steering Committee has also published data before it is uniformly available from all jurisdictions as a means to induce better data collection (SCRGSP 2010). Consequently, these reports have often encouraged the development of particular performance indicators.

Building evaluation skills and utilising academic expertise

Evaluation skills are in short supply within governments for various reasons including inadequate resources and low status being afforded to evaluation functions. The limited use of high-quality evaluations in Australia has also constrained the emergence of a pool of experienced and non-aligned private sector analysts and institutions — such as the US's highly-regarded MDRC (until 2003, the Manpower Demonstration Research Corporation) and Mathematica (PC 2010e). A preference by many agencies for in-house control of evaluations rather than drawing on independent external expertise has exacerbated these skill deficiencies in Australia.

Independent evaluation skills are unlikely to flourish where data is held in departmental 'silos'. Gruen and Goldbloom (2008) comment that 'microsimulation specialists pour into Nordic countries because of their liberal approach towards sharing statistics'. The former director of the Melbourne Institute has emphasised the value of Australia's new national testing arrangements that will provide comparable information on school performance, before cautioning that:

> Researchers may in future gain access to the underlying unit record data … The risk is that economists may be excluded from …such access and would thus have little incentive to learn the idiosyncrasies of unfamiliar datasets … (Sedgwick 2009, p. 2).

Restrictions on access to data were found by the Commission to have limited the emergence of a 'critical mass' of specialist gambling researchers in Australia, whose research might otherwise have helped identify effective policies to reduce problem gambling (PC 2010b). More generally, analysts have sometimes had to laboriously extract limited, aggregated data using software to 'discover' the underlying data points from deliberately restrictive official presentations of information (Leigh and Thompson 2008; Harding 2008).

A common rationale for not sharing data — especially micro data on the experiences of particular firms, individuals or families — is the need to protect privacy. This is a legitimate, but also a surmountable, concern. Techniques to

obscure data fields that might reveal identities are available. The ABS routinely uses such measures in making 'confidentialised' unit record files available.

External researchers are more likely to contribute to broadening Australia's evaluation skills base where data are accessible, evaluations are adequately resourced and results can be published. Better coordinated jurisdictional experimentation, including shared learning on implementation and administration, also provides opportunities to build public sector evaluation skills.

Meeting Australia's data needs

Policy formulation in Australia, especially in the human capital areas that are COAG's current focus, has been hampered by data limitations. Sometimes the necessary data have not been collected; sometimes the available data have limited applicability or are too partial for meaningful analysis; and sometimes data exist but are inaccessible. These problems are well recognised. The National Statistical Service initiative — a community of government agencies led by the ABS — is seeking to broaden the breadth of information that is supplied by statistical producers in order to better advance policy evaluation (NSS 2010).

Access to data improves analysis

Overseas experience demonstrates that greater access to micro data has helped to identify which social policies work best. Often the innovative use of data has been led by academics. At the Commission roundtable on 'strengthening evidence-based policy' held in August 2009, participants reported favourably on revised ABS pricing for data sought by academic institutions. Previously, Australian academics had to buy individual access to confidentialised record files. They could more readily analyse policies in the United Kingdom, United States and Canada — where data are freely accessible — than they could in their own country. Participants noted that better access to data usefully broadens the ranks of analysts beyond public servants, whose analysis is often confidential to governments and who are constrained from engaging in public debates (PC 2010e).

Making data public (subject to appropriate privacy protections) allows for independent verification of official evaluation findings, enables sensitivity analyses and experimental use of new methods, and encourages additional research of direct interest to government at little cost (Chapman 2010; Smith and Sweetman 2010). Allowing more analysts to corroborate or challenge official findings in turn strengthens the quality of analysis. An added benefit is the impetus provided by input from data users on how to improve the quality and usefulness of the data sets.

Longitudinal data is valuable

Access to comprehensive databases, especially longitudinal data, that track individuals' experiences over time, can promote policy relevant research on a wide range of issues. The usefulness of longitudinal data is exemplified by New Zealand's Dunedin study, which has followed 1000 individuals born in that city in 1972–73. The study has generated over 1000 reports, including research that helped frame policy responses to antisocial behaviour in New Zealand and in other countries (Scobie 2010).

The experience thus far with the innovative Household, Income and Labour Dynamics in Australia (HILDA) survey is further testimony to the valuable insights that can be gleaned from longitudinal data. It has been used to explore issues as diverse as: changes in household wealth; consequences of long working hours; credit card debt; dietary habits and health of people in different socio-economic groups; the effect of work-related training on earnings; and interactions between health, disability and specific medical conditions (MIAESR 2009, 2010). Other datasets — such as the Longitudinal Study of Australian Children, the Longitudinal Study of Indigenous Children, and the Medicine in Australia: Balancing Employment and Life survey — offer similar prospects of improving understanding of key policy areas.

Unlocking 'data silos'

As noted, evaluation skills are unlikely to flourish where data are closely held or in departmental 'silos'. Apart from reducing access to data, this can lead to wasteful and burdensome duplication. The Advisory Group on Reform of Australian Government Administration, established by the Prime Minister in 2008, has argued for breaking down 'silos' in administration and data collection:

> Advances in information technology are making a stronger relationship between citizens and government possible. …. The Blueprint recommends that the Australian Government become more open and that public sector data be more widely available, consistent with privacy and secrecy laws. (AGRAGA 2010).

The Department of Finance and Deregulation has a lead role in making public sector data more accessible. A vehicle for this aspiration is Government 2.0 — based on Web 2.0 collaborative tools — which aims to inculcate a culture that government information should be accessible by default in the absence of good reasons to the contrary. The Government 2.0 Taskforce (2009, p. 47) recognised that:

When information is released it creates new and powerful dynamics which can drive innovative use and reuse. Allowing the commercial, research and community sectors to add value to it can provide important social and community benefits.

The Taskforce recommended that, subject to privacy and confidentiality considerations, the Commonwealth make available the data it 'owns' and negotiate with parties to release shared or privately owned data. The Government has given its in principle agreement to this (Australian Government 2010).

The National Assessment Program for Literacy and Numeracy, together with the 'My School' initiative, exemplify the open provision of new bodies of micro data in education. COAG's National Education Agreement requires the Australian Curriculum Assessment and Reporting Authority to manage school assessment data and publish 'relevant, nationally comparable information on all schools' to allow performance comparisons of like schools (COAG 2008c). In endorsing the framework the (then) Minister stated:

> … lack of transparency both hides failure and helps us ignore it…And lack of transparency prevents us from identifying where greater effort and investment are needed. (Gillard 2008)

The notion that misunderstanding or misuse of data on school performance is a sufficient rationale for governments not to distribute that data widely is under challenge. While the misuse of data to advance particular agendas is an everyday problem with all types of data, transparency and greater familiarity with the data and their limitations, in concert with increasing examples of good evaluation, should lead to mature community management of those risks.

Improving data bases to achieve affordable gains

Standard reporting formats allow mapping of different data conventions to standardised definitions, simplifying the use of such data for policy analysis. Standard Business Reporting using XBRL (eXtensible Business Reporting Language) has been implemented for financial reporting in Australia leading to better, cheaper data and large savings in compliance costs.

Linking data bases provides a major opportunity to use existing data collections more effectively. For example:

- access to confidentialised unit records now means that taxation policy reviews need no longer be confined to assessing impacts on the 'average Australian' but through microsimulation modelling can now examine distributional impacts on finely graduated cohorts incorporating income, age, gender and marital status

- data matching in relation to social security benefits is a critical compliance tool akin to the Australian Taxation Office's access to combined records which allows it to assess tax declarations against expenditures (such as vehicle registration) and income sources (such as interest) to reduce tax evasion.

Generally, data linking opportunities have been achieved through policy collaborations and a desire to ensure program integrity through robust compliance mechanisms. However, the potential benefits for linked datasets extend beyond this.

An ability to combine data sets can deliver a boost to the evidence base by providing a richer source of information for research and evaluation. For example, New Zealand's Inland Revenue Department receives monthly returns filed by firms, listing all paid employees' earnings and tax. This is linked to the firm level data in the Longitudinal Business Frame, and also with benefit data from the Ministry of Social Development. The linked databases have enabled many policy issues to be studied, including: the effect of minimum wages on teenage employment; employment rates of former benefit recipients; labour productivity; implications of changes in workforce composition over the business cycle for labour productivity, job mobility and earnings dynamics (Scobie 2010, Stillman and Hyslop 2006).

Ultimately, it may be feasible to combine datasets to provide analysts with insights into people's lives over the lifecycle, from early childhood to retirement. Such information could constitute a powerful tool to understand why policies work or fail for particular groups of individuals. This is especially important in government policy areas where the discipline of competitive processes to drive reform is absent.

Planning for relevant data collections

Better data is more likely to be assembled at lower cost and with greatest benefit if collection of that data is part of a strategy focussed on policy needs. For example, in its recent inquiry report, the Commission found gambling policies needed to rest on clearer thinking about the nature of the problems gambling can cause, and evaluation of the relative net benefits of different policy options. It identified better ways of coordinating data and governing quality research (PC 2010b).

Participants in the Commission's study of the 'not-for-profit sector' identified benefits to the sector and to governments from a well-conceived framework for impact measurement, including achieving better public policy outcomes for funding that is directed through the sector (PC 2010d). The report recommended a structure by which the impacts of the sector could be measured and the most useful data on the sector could be built and used (box 1.3).

Box 1.3   **Improving the evidence base for the 'not-for-profit sector'**

A Commission study into the not-for-profit (NFP) sector found many NFP's struggle to meet government reporting and accountability requirements, which are often costly and not appropriate for all NFPs. The study identified paths to improve the evidence base on NFPs, including proportionate reporting requirements, more regular publication of the ABS's satellite accounts on the sector, and better data comparability across NFPs.

The Commission recommended that governments should:

- commit to basing reporting and evaluation requirements in service delivery contracts on a common measurement framework

- ensure that information generated through performance evaluations are returned to service providers to allow organisations to benchmark their performance

- establish a centre to promote, lodge and access 'best practice' evaluation and to support future 'meta analysis' of evaluations as they accumulate (PC 2010d).

Drawing on evidence from regulation benchmarking

Regulation reform is a fundamental means of achieving improved economic performance. A combination of good process and sound evidence and analysis can help drive improvements. Information collected by the Commission from its annual regulation benchmarking and regulatory burden studies has identified costly and unnecessary differences across Australian governments in regulatory practices, and inefficiencies in regulatory approaches.

This work has highlighted the importance of studying not only the formal regulations in place, but also how they are administered in practice. The work rests on the voluntary inputs of affected citizens and organisations, collected through consultation, submissions and surveys. The provision of such information has:

- helped signal emerging pressure points for reform, such as in aged care

- identified differences in enforcement practices and risk management — benchmarking of Australian and New Zealand food safety regulation found that Australia's regulatory system for exports relies less on electronic processing that could reduce compliance costs and is less able to accommodate outcome-based standards in the domestic food safety system than New Zealand's (PC 2009i).

- helped sustain progress and support for improvements such as in occupational health and safety regulations, where benchmarking state, territory and Commonwealth practices pointed to remaining problems in some jurisdictions that would be removed through implementing the Intergovernmental Agreement for Regulatory and Operational Reform in OHS (PC 2010g).

While the work to date has helped highlight areas requiring reform, the need for agreement across many jurisdictions can make progress difficult.

An evidence-based approach requires resources

Data collection and research can consume significant resources and time if done well. In a budget constrained environment, governments often perceive such activities to be expendable relative to their 'coalface' functions. However, such savings can prove illusory when weighed against the potentially greater and enduring costs of implementing polices that fail or that require substantial revision. Moreover, data acquisition and evaluation may be required later for audits and reporting obligations. Anticipating these needs and collecting data earlier for policy design and refinement can be worthwhile. Compliance costs from data collection can also be minimised if the principles of Standard Business Reporting, now used in the financial reporting sphere in Australia, can be applied more broadly. Those principles include:

- starting any data collection process by considering first what use can be made of data that business collects for its own purposes

- ensuring that, where possible, data is 'collected once — used often'.

Governments not only need to commit to better resourcing of evaluations, but research bureaux need to be able to operate with sufficient autonomy to pursue solid analysis of impacts under more than one policy option. Where evaluation units do exist within policy departments, they are sometimes constrained in the frankness of their (public) evaluations.

Evaluations by academics, independent consultants and private 'think tanks' potentially offer a remedy, but their utilisation in Australia has been relatively limited (PC 2010e). The Secretary of the Department of Prime Minister and Cabinet has recently stressed that the public service 'must work with people from the private and community sectors, think tanks, academics, stakeholders and members of the public. And we need to carve out time for thinkers within the APS to enable them to do long-term, creative work' (Moran 2009).

One example of such collaboration is the venture between the University of Melbourne and the Victorian Government that resulted in a new experimental economics laboratory in late 2007. The laboratory promotes experimental methods in economic research, assisting policy makers to understand how people's decisions are influenced in various situations and to design innovative policy processes, including iterative refinements before proceeding to field trials (see also PC 2008n).

The value of trials

The urgency that sometimes attends policy development and implementation can be self-defeating. Indeed, for national programs, the most expeditious and cost-effective path will often require evaluated trials. The cost of policy 'misfires' for a national rollout can be large. Relatively small investments in trialling policy reforms, sequentially rolling out policies to facilitate progressive improvement, and collecting baseline and other data can assist policy design and implementation, without adding to overall implementation times. There have been some positive recent developments.

• On 1 July 2010 the Australian Government commenced a 12 month trial for a new drought reform package in Western Australia, in partnership with that government. The $23 million pilot will test new measures to assist farmers to prepare for future challenges. After 12 months it will be assessed with a view to developing a new approach to be rolled out nationally.[1]

• As part of the Smarter Schools partnership agreements, Victoria will trial school- and teacher-based rewards over 2010–13. The 'Teacher Rewards model', which provides annual bonuses for top performing teachers, involves piloting 'two teacher pay bonus models at up to 75 selected Victorian government primary and secondary schools' (Victorian Government 2010, Pike 2009).

• The Australian Government aims to improve student outcomes by giving principals and school communities more control over how schools are run. Sequential roll-out for 1000 schools will commence over 2012–13. National rollout for most schools is to occur by 2018 'informed by an iterative evaluation of the first 1000 schools …' (Gillard 2010).

Making the most of these pilot programs will depend on the quality and transparency of evaluations, and subsequent policy actions that have proper regard for the results and lessons learned from the pilots.

Even simple methodologies can be revealing

Good evidence on how policies are working does not always necessitate extensive data, sophisticated quantitative techniques or the so-called 'gold standard' of randomised control trials. Often, qualitative analysis using simpler methodologies

---

[1] The (then) Minister for Agriculture, Fisheries and Forestry stated that 'there'll be challenges that we haven't fully anticipated. … I make no presumption that what we announced today will work perfectly. We want to … work it through in the best possible way.' (Burke 2010).

can shed light on policy performance. Disclosing information can, of itself, promote positive outcomes.

In the late 1980s, the Department of Health in New York began collecting information on every hospital patient receiving heart bypass surgery in the state. The publicly reported information identified hospitals and the patient's outcome. Citing this case, the Treasurer contended:

> … between 1989 and 1992, mortality rates for cardiac operations in New York State hospitals declined by over 40 per cent state-wide … It happened because hospitals and surgeons didn't want to be labelled as the worst in the State and the public reporting led to improvements to their cardiac surgery programs. (Swan 2008).

Other examples from the health field achieved through collating basic data and interpreting available information include the substantial reductions in deaths from Sudden Infant Death Syndrome (SIDS) and reductions in revision surgery for joint replacements (box 1.4).

Box 1.4   Better health outcomes through general analysis

The recommendations to parents on how to minimise the risk of SIDS have reduced Australian deaths from SIDS by more than 80 per cent. This illustrates a successful intervention that rested on a mixture of epidemiological studies, pathology studies and case studies, rather than on quasi-experimental methodologies. Rogers (2010) provides an outline of tools that may be appropriate to evaluating policy issues of differing degrees of complexity. Examples include 'general elimination methodology' and 'multiple lines and levels of evidence'.

The risk of revision surgery following hip and knee replacements in Australia (20-25 per cent) is comparable to most other countries, but higher than in Sweden (10 per cent), owing to the impact of that country's long-standing hip and knee registries. An Australian National Joint Replacement Registry (NJRR), which became operational in 2002, collects data on all joint replacements and the incidence of revision surgery. One reason for Australia's higher failure rate was the use of multiple types of prostheses. The NJRR data showed that newer, generally more expensive, prostheses did not always deliver better outcomes (Graves and Wells 2006). The NJRR provides evidence which surgeons can use to reduce the risk of surgical revision. By 2006, this information had reduced joint replacement revision operations by 1200 annually, benefiting patients and saving between $16-$30 million per annum (ACHR 2006).

Sources: Rogers (2010); Graves and Wells (2006); ACHR 2006.

The Commission also has long used qualitative analysis and simpler methodologies where appropriate when the available evidence base cannot support more advanced analytical techniques. For example, industry assistance reviews have traditionally

drawn on partial measures such as estimates of effective rates of assistance and consumer tax equivalents. Other examples include the Commission's work on:

- the implications for Australia of firms locating offshore — 150 mining, manufacturing and service firms were surveyed to determine the domestic policy influences considered important to their decision to invest offshore (IC 1996).

- the impact of an outbreak of foot and mouth disease — outbreak scenarios were assessed by developing a cash-flow model to estimate trade and production effects, supplemented by general equilibrium modelling to estimate economy-wide impacts (PC 2002c)

- the market for retail tenancy leases — like many inquiries, this proceeded on the basis of public hearings, submissions, industry data and lessons from jurisdictional variations in lease registration and dispute resolution (PC 2008o).

Complex policy questions can benefit from sophisticated analysis

Many complex social and economic issues can only be addressed properly using analytical techniques that disentangle various potential influences. Does the New York hospitals experience (above), for example, prove that information provision led to competition between hospitals that saved lives, or did new medical technologies or changes in case mix play a part? Some evidence comes from the United Kingdom, where in 2006, the Government introduced a policy to give patients a choice between hospitals as well as information on the quality and timeliness of care. Prices remained set centrally. Gaynor et al. (2010), using a quasi-experimental research design, found that informed patient choice saved lives without raising costs. Their research design allowed them to rule out that the improvements were driven by other factors such as hospital case mix differences or patients' socio-economic status.

Establishing the *causal* effects of government programs or policies can be very difficult in areas such as health, aged care, education, and Indigenous disadvantage. Investigation in such areas would generally benefit from better data and evaluative impact assessment methodologies that: identify the counterfactual against which a policy's impact is being addressed; address problems of multiple causation; and avoid biases that can bedevil some simpler analytical approaches.

Analytical methodologies have emerged in recent decades that allow analysis of social and economic policies to separate the impact of a particular policy from other influences. There is growing experience with social and economic applications of randomised controlled trials, and of econometric methods for 'quasi-experiments' such as instrumental variable approaches, differences-in-differences analysis and

regression discontinuity (Smith and Sweetman 2010; Leigh 2010; Angrist and Pischke 2010; Imbens and Wooldridge 2009).

The benefits of applying such analytical techniques are exemplified by some notable international successes in high quality evidence-based evaluations of complex social policy reforms (box 1.5). These include:

- the 'Progresa' welfare reform in Mexico in the 1990s (renamed 'Oportunidades' in 2002) which used a large trial followed by an evaluation prior to wide scale program roll-out. 'Progresa is why 30 countries worldwide have conditional cash transfer programs' (Gertler, in Angrist and Pischke 2010, p. 4). Indeed, Progresa influenced Australia's current remote communities conditional transfer trial which involves obligations on recipients to, among other things, ensure that children attend school and are kept safe from harm (CYI 2007, 2010).

- welfare reform in the United States in 1988 and 1996, the foundations for which were built on extensive earlier state-level experimentation, which analysed the impacts of financial incentives, program assistance, and compliance strategies through measured outcomes up to five years after the program intervention.

Microsimulation modelling has been used in Australia for some time (for example, the National Centre for Social and Economic Modelling's suite of models and Treasury's retirement income models). Nevertheless, Australia has been a limited user of newer analytical advances (Leigh 2010). This is notwithstanding that the capacity to support the application of sophisticated methodologies has improved considerably owing to technological advances.

- Data capture techniques have improved. This is exemplified in wholesale and retail trade where barcodes and sensors enable vast data capture in real time (Johnston et al. 2000). However, data collection for social programs remains resource intensive owing to the complexity of programs, the multiplicity of transactions and the need to match data to socio-economic information.

- Data storage is now cheaper, permitting large collections of microdata of individual, household or firms to be accessed. For example, Amazon.com's two largest databases are said to hold 42 000 gigabytes of data; storage that would have cost over $30 billion twenty years ago (Gruen and Goldbloom 2008)

- Computing power has been vastly enhanced, enabling analysis of large databases and linkages among databases. Modelling the impact of rising carbon prices on the Australian economy over the next 100 years now takes around 10 hours on a desktop computer, whereas two decades ago such computations would have taken over a year (Gruen and Goldbloom 2008).

Mexico's Oportunidades (Progresa) program

In 1995 Finance Ministry officials proposed replacing in kind distributions of milk and tortillas, and subsidised bread and tortillas, with targeted cash transfers to mothers contingent on household members attending health clinics and children attending school. The ambition was for one comprehensive transfer program for poor households to improve children's education, health, and nutrition. The proposition was met with questions about its efficacy: Would cash transfers lead to more spending on tobacco and alcohol? Would giving cash to mothers lead to family disruption and violence? Would making cash transfers contingent on outcomes be operationally feasible?

To answer these questions a pilot involving 31 000 households was implemented. The subsequent evaluation of the controlled experiment was based on advanced statistical techniques and it found that cash transfers did not promote inappropriate use of funds — most households valued the link to health services — or family disruption. However, it highlighted that a full scale rollout would require revised targeting and better data collection. Moreover, while the pilot showed that a large scale program could yield substantial benefits with low risk, it also illuminated government agencies' unwillingness to coordinate a large-scale operation.

Lessons from the pilot enabled the program to be tailored to address these operational issues. Further public evaluation demonstrated the program's success which endowed it with popular support, insulating it from attack in election campaigns and a subsequent change of government (Levy 2006).

The role of state trials in the evolution of welfare reforms in the United States

The foundation of US social welfare reforms lay in a period of state-level experimentation beginning in the 1970s and spread over about 15 years in some 40 states. A federal law allowed experimentation and states conducted trials of ways to increase workforce participation by those on welfare, which led to a large body of high-quality evaluations to identify the most effective approaches.

These trials also gave state bureaucracies the experience and confidence to administer larger scale reforms. Moreover, successful state experiments created a constituency of state and federal politicians prepared to support national reform: the Federal government was not trying to persuade or 'pay' states to make reforms they were not otherwise motivated to make (Haskins 2010).

Sources: Levy (2006); Haskins (2010); PC (2010f).

Such developments support the use of more powerful analytical methodologies. For example, the Commission applied experimental 'multivariate analysis' — a statistical technique in which two or more variables are analysed simultaneously — to assess the relative efficiency of public and private hospitals (PC 2009d).

Supplementary modelling drawing on additional years of data found that hospitals are operating around 10 per cent below best practice (PC 2010h).

Importantly, the stock of evidence-based evaluations from overseas can sometimes provide lessons for Australian policy analysis, provided appropriate allowance is made for country-specific differences. Similarly, there are opportunities to draw on experiences in other policy areas where there is a prospect of some transferability of policy action.

Institutional support for better analysis

Good policy processes and effective institutional arrangements can support quality information gathering and analysis, and help ensure they bring about policy improvements. But there are many reasons why this ideal is often not met. Short electoral cycles militate against investment in quality evaluation and towards limited evaluations designed to show short-term policy pay-offs. There are also political economy forces at play. In federations, the costs of collecting data and evaluating policy in a state or territory usually fall only on that jurisdiction. While most of the benefits and the political risks of identifying poorly-performing policies also accrue to that jurisdiction, some of the benefits of learning from policy successes and failures can accrue to every jurisdiction and all Australians.

This tension between the internal costs and external benefits of evaluation has been recognised as having caused an 'evaluation gap' in analysis of international aid effectiveness (EGWG 2006). There are several international innovations that attempt to capture the 'external benefits' and incorporate them in the decisions to finance, undertake and share the results of good data and evaluation. Most of these involve types of 'evaluation club', typically with commitments to: principles of rigorous, transparent evaluation; sharing lessons learned; improving the standard of evaluation; and sometimes, helping to fund high-quality evaluations (box 1.6).

Support for high quality evaluations such as those provided by the US Office of Management and Budget or the Coalition for Evidence-Based Policy has not emerged in Australia. There are some 'policy hubs' operating at various tertiary institutions (for example, the Melbourne Institute) and evaluation practitioners share their insights in groups such as the Australasian Evaluation Society, whose 1000 members include evaluation practitioners, managers, teachers and students of evaluation (AES 2010). The Society promotes professional standards of evaluation, but works more with the methodology of good evaluation, than with the policy processes of funding and using evidence well. But there is scope to do more.

The Council of Australian Governments is a key forum

The key institutional support mechanisms that could be said to address the 'evaluation gap' in Australia arise through federally-based machinery such as Ministerial councils, steering committees — such as for the review of government services provision — and intergovernmental agreements.

The NCP involved governments signing intergovernmental agreements to which rewards for achieving particular reform milestones were attached. This approach continues in the COAG reform agenda, which uses a blend of cooperation and competition through National Partnership Agreements (NPAs), with 'inducements' from the Australian Government through associated National Partnership payments (NPPs) and specific purpose payments.

Detailed policy design remains the province of the states and territories, but working groups devise frameworks of indicators to illuminate progress towards outcomes, objectives and targets. The COAG Reform Council (CRC) publishes the

output and outcome performance indicators to highlight differences among states and territories and facilitate performance improvements (COAG 2008h, 2010b). This provides potential to identify jurisdictions with higher relative performance based on good practice. This is very important for the human capital arena where government spending is substantial, as are the policy challenges (box 1.7).

Box 1.7   **Human capital reform — its significance and complexity**

Commonwealth, state and territory government spending on education, health, and social security and welfare is substantial — it totals over 60 per cent of all general government expenditure. Spending more effectively in these areas could pay big dividends by reducing waste and achieving superior outcomes for given expenditures.

Getting the best value from human capital reforms will require strong evidence because these areas are complex and politically sensitive. There are multiple causes for observed outcomes, and alternative pathways to address those outcomes that are seen as sub-optimal. For example, childhood obesity may be a significant risk factor for type 2 diabetes and cardiovascular disease in later life, but there are genetic and lifestyle influences too. Individual and family choices can also thwart policy intentions. Choosing the most efficient mix of interventions that is most likely to reduce later disease requires good evidence.

Moreover, human capital reforms may not be well suited to 'one size fits all' solutions. For example, nearly one third of Indigenous children in Year 3 have skills below the national minimum standard, compared to 6 per cent of non-Indigenous children (CRC 2010). But achieving better educational outcomes for Indigenous children will need to address multiple causes of disadvantage. Similarly, while homelessness can be approached from the perspectives of housing shortages and poverty, mental illness is a strong contributor.

Sources: ABS (2010); PC (2006b); CRC (2010).

The performance-linked inducements through NPPs and the consequential monitoring of performance aim to create incentives for jurisdictions to apply themselves to achieving COAG's national reform goals. Ideally, this should encourage an evidence-based approach in order for states and territories to determine what works. For example, under the renegotiated National Partnership on remote Indigenous housing (COAG 2009e), payments for 2010-11 and 2011-12 were recently increased for Western Australia and decreased for Queensland and South Australia to reflect their relative performance in meeting 2009-10 targets for completing new, and refurbishing existing, houses (Macklin 2010b).

That said, while rewarding better relative performance can promote policy learning across the federation, this need not mean that optimal, or even cost-effective, policies will end up being pursued — 'less bad' policy initiatives can be rewarded

and, through this, even locked-in as the national best practice norm. Moreover, the indicators approach presumes causality from policy inputs to the exclusion of other, possibly significant, factors. They also may include 'noise' in the form of specific geographic or jurisdictional contextual differences that make benchmarking challenging. These are reasons why, for example, the Commission complemented the available partial indicators with multivariate analysis in its study into public and private hospitals (PC 2009d, 2010h).

This challenge of assessing program performance and effectiveness has been recognised by COAG, which has charged the Commission with reporting every two to three years on the impacts and benefits of COAG's reform agenda, to complement the CRC's monitoring role. The Commission is to assess the economic impacts and benefits of realised COAG reforms, whether Australia's reform potential is being achieved, and the opportunities for improvement.

With COAG processes now identifying and rewarding successful policy performance through the NPPs, Australia is well placed to trial some of the collaborative approaches discussed above to strengthen the evaluation of policy innovation. NPPs offer the opportunity to apply stronger analytical tools to finding what policies work best and in which circumstances, thereby avoiding resources being committed to policy changes that may be sub-optimal.

Further progress

As noted, during the 'reform era' of the 1980s and 1990s, Australia earned a reputation internationally for a policy approach based on independent advice, interrogation of evidence, consultation, and building constituencies to support structural reforms (OECD 2005). This paid dividends, with those reforms lifting productivity and living standards substantially, reversing a secular decline in Australia's international ranking.

The need for evidence-based public policy remains as important today. Evidence and good process can secure better regulation, more efficient infrastructure investments and, critically, more cost-effective social policy. Because human capital policies inevitably must deal with individuals with differing socio-economic, demographic, regional and cultural characteristics, they cannot be hastily conceived. Basing public policy on sketchy data or quick surveys (or worse, focus groups) that may involve people recalling past behaviours or predicting their responses is fraught. Moreover, the technological and methodological capacities to progress effective evidence-based policy have advanced considerably.

A number of countries have led the way in addressing these problems. They include not only advanced economies such as the United States, with a rich history of quasi experimentation to test the efficacy of social programs, but also middle-income countries such as Mexico. Australia lags in such endeavours.

An action agenda for evidence-based policy

There are several priority actions needed to develop the evidence base, support better analysis of evidence, and help build or sustain community support for necessary policy reforms. There is a particular need to:

- encourage a culture of open access to affordable data (subject to confidentiality protections) for government users, academics and other researchers

- promote early planning to ensure that data collection is relevant and evaluations can reliably assess policy impacts relative to 'baselines'

- facilitate standardisation of reporting formats and greater linking of datasets to enhance the available evidence base

- ensure research agencies and evaluation units within government departments have sufficient time and resources to undertake or commission quality research

- entrench requirements for public consultation — not only to gather evidence, but also to test competing views and expose draft policy proposals to scrutiny

- ensure that review processes are independent and transparent — especially on contentious matters where judgment is called for

- encourage policy trials and experiments that involve rigorous assessments, rather than relying on the anecdotal experiences of selected stakeholders

- investigate iterative policy implementation approaches where there is significant uncertainty or large gaps in the evidence

- pursue institutional experiments such as 'evaluation clubs' to build analytical and evaluative skills, fund proper evaluations and share results

- ensure that the COAG model encourages 'best', rather than merely 'better', practice by linking some funding to requirements for jurisdictions to have ex ante evaluation strategies and develop policy evaluation skills.

Such actions will obviously involve a greater call on time and resources, and may run up against budget constraints and a political need for early action. But given that the cost of policy 'misfires' can be substantial and enduring, both economically and politically, the most cost-effective and sustainable path may ultimately rest on a slower, but better informed, start.