



OFFICE OF THE PRIME MINISTER'S CHIEF SCIENCE ADVISOR

Professor Sir Peter Gluckman, ONZ KNZM FRSNZ FMedSci FRS
Chief Science Advisor

Keynote address

IGPS (Institute for Governance and Policy Studies) Symposium “Identifying What Works: Using Randomised Control Trials in Public Policy”

7th July, 2015

Victoria University, Wellington

Evidence informed policy formation and implementation – an evolving landscape?

Thank you for the invitation to open this meeting.

I need not tell this audience how complex policy formation really is. It is not the idealised and rather simplistic cycle that is so often used to describe the process. What I want to focus on is the reality of how three sets of stakeholders interplay in the relationship between evidence, policy formation and implementation: these stakeholders are the scientific community (and in this talk I use a broad definition of the scientific community to include the social sciences as well as the natural science and engineering communities), the policy community and the political decision makers.

Each of these groups of stakeholders generally approaches policy making from their own perspective, their own set of needs and expectations and – importantly but rarely discussed – with potentially considerable hubris. So these relationships need to be parsed out gently.

Part of the challenge is to find a common understanding of evidence and how it should be assessed. Different stakeholders will have different views of what is evidence, how it is sourced and whether it is sufficient to support one course of action over any other. At one end of the spectrum, much of the science community would rely on the very formal processes of science to formulate a conclusion on evidence. In some cases, scientific expert opinion based on mastery of domain specific scientific literature could be offered as evidence, but in a number of areas of policy, the randomized controlled trial (RCTs) and meta- analysis of the empirical evidence are also gaining support - something that we shall return to. But at the other end of the spectrum is the political power of anecdote, but as has been said so many times the plural of anecdote is not data.

Globally there is a sense that evidence should play a greater role in the policy process than it has, not only in relatively linear matters but also particularly when it comes to those wicked questions – or what has been called ‘post-normal’ science - which dominate in the policy process. These are the issues where there is great complexity, knowledge is necessarily incomplete, there is a high public interest and the public values underpinning that interest are almost always in dispute. In turn, all of this means that there is inevitably a political dimension, including the fact that the need for decisions is seen as urgent. Climate change

is an obvious example but so are many other decisions regarding the relationship between the environment and economic growth, much of social policy has this characteristic and so has a topic like obesity. It is because of these characteristics that a generation of scholarship in the field of Science and Technology Studies has pointed out that evidence and policy formation have a much more nuanced relationship than seeing evidence as a purely technocratic input into the policy process. Indeed, there will always be questions of: 'whose' evidence backs policy decisions? and what were the processes by which it was generated? In highly contentious areas, even the most robust science will be subject to scrutiny. Transparency is paramount.

Indeed, why *should* scientific evidence have any claim to privilege or a priority space in policy formation? This is based on the claim that science is an epistemology that strives to be relatively values-free. Certainly the processes of good science (and not all science is necessarily good science) are defined in terms of removing bias and values from the collection and analysis of data but there are nevertheless values within the scientific process. The questions of what to study, what methodology and analytical framework to apply all require values judgments. But perhaps most important of all in a post-normal setting, is the question of the sufficiency of evidence and the inferential gap between what is known and what is concluded in order to advise action. In its most obvious form this is the danger of cherry-picking data or setting experiments to suit a position or lead to a particular outcome. Here, scientists can be as influenced by cognitive biases as anyone else, and it is neutralising these biases which is precisely what the scientific process is designed to address, in a formal sense. Of course the *application* of evidence is always going to be a values-rich exercise, and this is clear within the context of what you are addressing today. These issues are in no small part why there is a growing interest in the interface between the cultures and practice of science and policy formation.

The importance of parsing the understandings and contentiousness of evidence becomes clear when one considers the ways in which science-based evidence impacts on the policy process. Roger Pielke developed a somewhat idealised but rather helpful heuristic that classifies the types of scientist and use of science in the policy spaces. Two of the major types in Pielke's classification of interactions are: the issues advocate and the honest broker. The former is when the science is put forward to deliberately advance a position – this may and should be done with great validity to the science and on a sound basis but sometimes the arguments can be advanced with variable levels of fidelity to the data and there are times when scientists will advance a position claimed to be based on science but in reality this primarily based on advocacy and conviction. This may work in a political framing, but for science to have impact directly on policy formation, it is the honest brokerage approach that is desirable certainly for those in formal advisory or policy roles. The honest broker is distinguished by their attempt to describe what is known, what is not known about an issue, and the options that emerge – accepting that it is role of the policy and political process to add many other dimensions in reaching a policy choice. And here is the critical role of boundary structures for science advising - either in the form of individuals like myself, or through deliberative processes such as academy panels.

Science advisors are able to interact with the policy process at any stage and often iteratively, whereas deliberative advice is slower to develop and depends on the question being framed appropriately. One of the challenges is to ensure that when effort is put into deliberative advice there is a marriage between demand and supply. Too often academies

produce reports with great deliberation for which the policy demand is weak. One of the challenges is that the processes to develop academy reports must be seen to protect the integrity of the Academy and as a result, the timeframe for report development in the academy will rarely fit with policy development needs, and there can be different understandings of what is needed on the demand side between the policy maker and the academic. That is where boundary structures can assist the policy process in framing the questions appropriately and we have seen this approach used recently in NZ in the context of the fluoridation and asbestos issues where my Office engaged with the Royal Society of NZ to frame and produce joint reports.

The reality is that in a democracy, policy formation and political decision-making are and should be based on more than scientific advice alone. Science alone cannot decide whether or not a society should accept a particular trade-off between economic growth and environmental protection. But science can and should certainly inform the choices that society makes. But there are so many other inputs into policy making which have a dominant values dimension: affordability, political priority, ideological framework, electoral contract etc.

What science can do is provide a base for a range of options - based on what we know, what we do not know, estimates of impact and benefit, estimates of probabilities like spill-over effects etc. All of these questions are relevant to your meeting today. But the other dimensions of policy formation that I have mentioned are not something in which the scientist claims any expertise or privilege. This is why I prefer the terminology of *evidence-informed* rather than *evidence-based* policy formation. And I note that it is gaining currency.

But one place where science can play a much greater role and particularly assist the policy maker, and indeed the politician, is in developing greater insights and evidence about how citizens and users of services might respond to any particular option. There are many facets to how this can be achieved including behavioural insights, design methods that I shall return to, behavioural insights and the use of controlled trials and so forth.

The reality is that other considerations will often trump science in policy making. Consider two issues of recent and current controversy – climate change and genetic modification– both of which are matters of broad scientific consensus. Ironically, public advocates of action on the former are often advocates of inaction on the latter and vice versa, despite the respective positions of scientific consensus. Yet both sets of advocates will invoke science to support one position and reject science on the other. Why is this the case? Well there is considerable recent research to suggest that cognitive biases influence how we all look at evidence and filter information in accordance with how we want to be perceived. This means that simply piling on more scientific evidence does not change minds.

What is the solution? The first step, as Daniel Sarewitz recently reminded us, is to acknowledge that society not science must decide on such issues. The second step is to acknowledge the limits of science. Third is to do far better in publically explaining how scientific consensus emerges – it must never be confused with the conventional understanding of consensus. Fourth is, in my view, the need to acknowledge the complex interface between science, values and policy and therefore place science appropriately and transparently into the public, policy and political process. Sometimes this can be difficult given that extreme advocacy and rhetoric can limit informed public engagement.

Globally there has been a massive rise and interest in the use of science and evidence to inform policy formation and implementation. It is just over 50 years since the USA and UK appointed their first science advisors. Both countries have evolved different but strong approaches to both deliberative and informal advice. In recent years we have seen strengthening of science advisory systems in many countries but it was only last year that the first global meeting on science advice to governments took place. Over 40 countries were represented at this meeting in Auckland last year. I chair the planning group for the network of advice practitioners, policy makers and academics that emerged from that meeting.

In NZ it is 6 years since I was first appointed and my appointment is distinctive in both its level of independence and in the fact that my terms of reference give primacy to the use of evidence for policy formation rather than to proposing policy for the science system.

At the same time there is the Policy Project underway: this is a considerable effort underway to improve policy capability and the quality of policy – it is led by DPMC and I see these two developments as intertwined.

Over these last years, the Government has responded to my recommendations and departmental science advisors have been appointed in a number of core ministries with several more in progress. Their primary role is to improve and assist in the interpretation and use of evidence in ministerial policy development and to enhance the quality and appropriate use of internal and externally commissioned research. In some agencies, such as MPI, the role will become critical if a crisis occurs. The departmental science advisors also have important boundary roles in translation between the cultures of science and policy.

I chair regular meetings of this group of departmental science advisors, which serve both for informal mutual support but also to work with officials on specific matters. Matters that this group has discussed range from the use of big data in a variety of settings to public understanding of risk to specific technical matters where input has been requested. In this group, the DSAs are regularly joined by the chief economist and the chief statistician. The Deputy CE of the SSC and the President of the Royal Society join us for a number of these discussions.

Earlier this year the Ministers of Finance and State Services asked this group whether it would be able to provide independent advice on a number of social sector budget bids. This was a pilot scientific assessment process and we focused on 4 dimensions: The evidence for need; the evidence suggesting the likely impact of the proposed approach; systematic consideration of alternate and current approaches; how outcomes (as distinct from outputs or inputs) could be assessed and how the programme would be implemented in a way so it could be assessed. The level of interrogation was deep. In my reporting to the social sector ministers on behalf of the group, I explicitly acknowledged the other dimensions of policy formation and political decision-making. To quote

“The Science Panel has restricted its advice to the assessment of the evidential base supporting the bids submitted to it. We have not commented on the fiscal, political, societal or other reasons for either funding or not funding a proposed initiative, accepting that these are

properly matters for the political and policy process. The panel also understood that policy decisions must inevitably be made with an incomplete evidential base.”

From our viewing of a number of budget bids, there is no doubt that there was a highly variable quality of evidence presentation with a number of bids really not seeing the need for framing evidence clearly under these distinct headings. This will be addressed in the briefings to departments about future budget bids. In debriefings with Ministers and with Treasury officials it has been agreed that the science advisors will assist departments in making their bids more evidentially robust at a generic level and then will again have an ongoing role in evaluating the evidence base put forward in social sector budget bids. The processes are still being detailed, but lessons from the pilot have been instructive.

Up to this point, I have outlined both the challenges and the democratic need to bring science into the policy context. I have suggested both the theoretical and operational means to do this and provided an example of the types of approaches we have undertaken in NZ. But this meeting is about a specific approach to our common goal of better science-informed public policy: the use of randomized controlled trials (RCTs) in relation to policy making. So let me reflect on this topic in the light of my preceding comments.

First let me say that no one method can cover every circumstance and I am somewhat worried that an arrogance might emerge as to how one method might trump others. It depends on the circumstances, the question and the reality of context. Policy makers need to be aware of a range of analytical and methodological approaches and interpret and develop their options accordingly.

One can see the appeal of applying RCTs to the policy context where appropriate and where it is realistic – particularly as the use of robust and transparent science that is bespoke to the policy question, has the potential to resolve criticisms directed at either the expertise of individuals, or the application of more generalised evidence to a particularly contentious context. In general RCTs will have most attraction for relative specific operational questions, where confounders can be controlled for and the intervention defined. Indeed most examples of controlled trials in public policy have operated at this scale.

If we consider the assessment parameters set out by our science advisory panel (need, impact, relativity and assessment of outcomes), what RCTs can do, at least in theory, is answer questions about effect and effect size (impact) provided there is clarity as to the outcome being tested. This, in turn, leads to the vexed question of proxy measures where the outcome is too long into the future to be easily studied. Even in medicine, which pioneered RCTs, the latter is a very problematic issue. Designed well, RCTs can also answer questions about spill-over benefits and costs. Depending on how the trial is set up it can answer questions about comparing one approach versus another. These are the very questions that framed our assessment of budget bids.

But there are of course a number of challenges, theoretical and practical limitations. While it is easy to say RCTs are the gold standard, they do have limitations even in medicine. For instance, they depend on their interpretation, on the quality of the trial design, randomization and the matching of groups, compliance of research participants, and on the sample size and thus power of the study. Even in medicine we have seen many examples where these criteria have been variously interpreted as less than optimal and quite variable

results and interpretations are possible. Even when the results of different trials are combined through meta-analysis, very variable conclusions can be reached depending on the inevitably values-based judgments of inclusion and exclusion criteria. Hence there is always a need to see compatibility through multiple strands of evidence.

Another big issue is that RCTs are generally designed to address the average responses rather than to address variance; yet in much of what we want in developing social policy we are interested in the outlier, the tail and the distribution – for they are likely to be the points of policy focus.

The debate over the use of RCTs for policy in the social sector and often also in the health sector has been complicated by the debate over ethics. Is it ethical in a situation where normative arguments are strong to delay offering a programme to part of the population while the programme elements are assessed through RCT? But while from the policy and political process, such normative arguments are seductive, there needs to be an honest assessment of how robust that normative argument is. Often there is an exaggeration of certainty from normative arguments. There have been more than one example where relying on such arguments have led to adverse rather than better outcomes – putting babies to sleep on their tummies was a normative and scientifically unsupported argument and it was only from careful epidemiology that we learnt that putting babies to sleep on their backs dramatically reduced the incidence of cot death. Driver education at high schools is another example where normative arguments were shown not to be supported by the evidence. The example of steroid treatment for head injury is presumably well known to all of you as it is used in the influential paper of Laura Haynes, Ben Goldacre and others “Test, Adapt and Learn” put out by the Behavioural Insights team in the UK.

But controlled trials are not always possible – for example the bulk of evidence supporting fluoride in water comes from observational studies for obvious reasons and attempts to use RCTs and meta-analysis to address that question have had a number of limitations. Depending on the interventions, controlled trials may be compromised by the so-called control group changing their behaviour through spill-over knowledge or influence, or the background rate of the problem changing. Again this experience is not uncommon in medicine.

Despite the challenges, I can envisage at least two distinct ways to operationalise trials to inform policy:

The first is where the trials arise from commissioned studies whether academically initiated or policy sector initiated. A question is put, a target population is identified and randomized explicitly (and the consent issues here depend on the nature of the intervention) and the study conducted and analysed. Such studies are in many ways forerunners of the policy process and the RCT must be seen in this context as one more input into the evidence mix, albeit it with particular weight – assuming the trial is done well and the population is appropriate.

The second method is more directly relevant in the policy process but has been used inadequately in my opinion even though it may not have all the formal reliability of a very formal RCT. Simplistically it may be through the well-designed and controlled pilot study or the use of programme introduction itself as a pseudo-trial. We need to use pilot studies

much more extensively but when we do such pilots, we need to make sure we understand the baseline condition and have properly defined the control group. Too often I have seen a failure to understand the importance of baseline data for both the control and test groups to be sure they are comparable or that confounders can be addressed. Indeed, too often a programme has been introduced in a context of scientific uncertainty, yet it is not assessable because no baseline data was collected. Here we have policy and political hubris leading to an unwillingness to adequately invest in implementation science (sometimes called intervention research). Sometimes this is because decision makers do not want to admit that something introduced may not turn out to have worked, or where they want to appear more decisive about introducing a programme than is justified. In some cases, there is the pervasive assumption that pilot studies waste money and that we should just 'get on with it'. But the experience of the youth mental health programmes introduced by this government shows that the public can see the wisdom of accepting that in complex areas things have to be tried and assessed and can accept political statements of uncertainty when a new programme is introduced – indeed I think that honesty and measured action is welcomed by the public.

How often have we heard “we know this will work”, when in fact a decade later it turns out that we have no idea what works and what does not. My bias and I suspect that of many politicians and commentators is that we invest in a lot of programmes that do not work but we do not have a way of knowing or stopping them. Indeed a major reason the Social Policy and Evaluation Research Unit SuPERU has been set up, now chaired by Len Cook, has been to enhance the quality of implementation science to inform the social sector. This in turn must lead to a more self-learning policy process. It is no accident that we have made early appointments of quantitative social scientists as department science advisors in education and social development.

I think an under-valued approach is the more strategic use of programme implementation itself as a form of pseudo-trial. If good baseline data is collected, then we have an important start. But if we think about the reality that so many programmes cannot be introduced at scale nationally at once, then there is an inherent ability to monitor and evaluate provided confounders are adequately considered. This needs to be considered at the outset of implementation as some form of randomization is needed. It can also allow for more appropriate mixed methods than a classical RCT might, where qualitative and quantitative data can be collected and integrated throughout pilot implementation so that there is no risk of strictly technocratic outcomes (an important consideration in the social sector especially). However, the short political cycle we seem to be stuck with and the consequent political considerations can interfere.

But governments are realizing that within constrained budgets there may be many programmes that are either not achieving, could be better targeted or have spill-over benefits or costs, that better decision making will lead to better outcomes. Most governments want to make better investment choices and seem to understand that even within ideological constraints of different political parties, evidence-informed decision-making has ultimate electoral benefit.

Public policy has a range of analytical and methodological tools to use that are science based. They range from data analytics, to RCTs, and to post-hoc evaluation. These and other approaches all have their place, just as RCTs are not the only approach to advancing

clinical medicine. But expert judgment is needed both to evaluate the methodologies, the data and to opine of the sufficiency of evidence.

In the end, evidence alone does not make policy, but better policy will come from better use of evidence. Indeed the investment approach now being advanced by the Government through Treasury and related agencies requires evidence for wiser investment and we can all assist in that goal.

Thank you.