# Institute for Policy Science Quality Control: Application to Great Barrier Reef Science.

Dr Peter Ridd
Independent Scientist
34 Mango Avenue, Mundingburra, Townsville
peterridd@yahoo.com.au
049790035

## Executive Summary

**The Problem:** Some of the science upon which governments base expensive decisions is not as reliable as it needs to be. This problem is particularly acute for the science of the Great Barrier Reef where it is likely that some of the funds to save the reef will not be spent on the most urgent environmental problems. It is also possible that some legislation that is based on questionable science will result in little environmental benefit but will cause significant costs to industry.

**The Solution:**

Set aside 1% of the recently announced $500 million "Reef Rescue" funds to set up an "Institute of Policy Science Quality Control" that would do truly independent checks on GBR Science.

Allocate 5% of the $2 billion per year funding for the Australian Research Council and National Health and Medical Research Council to checking, testing, and replicating "Policy-Science"

**The Cost:** Nil. Funds are already allocated.

It should be noted that the Dutch and US equivalents of the ARC, (Royal Netherlands Academy of Arts and Sciences, and the National Science Foundation) have both recently instituted policies to do science checking and replication studies.

## Extended Summary

There is growing evidence that some of the scientific results upon which large public policy decisions are based, have serious flaws due to problems with the quality control methods being used in many scientific institutions. This proposal seeks to improve the reliability of the science upon which large public policy decisions are based (defined as Policy-Science) with a specific proposal for the science underpinning the plans to save the Great Barrier Reef (GBR). It should be noted that;

- Peer review is the primary Quality Control (QC) system for most Policy-Science. Although most of the public believe that peer review is an exhaustive process involving many scientists spending perhaps months checking results, it is in fact a very cursory process that may take a couple of reviewers only a few hours to complete.

- Checks of peer reviewed work regularly find that around half is flawed. This remarkable fact, which has only come to light over the last decade, has become a major cause for concern in the mainstream scientific community. It is now called "The Replication Crisis".
- Far more rigorous QC processes than peer review should be used for policy decisions worth many millions or billions of dollars.
- Unlike governments, industry rarely relies solely upon peer review for expensive decisions due to its unreliability.

In the next few years the Federal Government will spend around $500 million to save the GBR. It will be important to spend this money as efficiently as possible to maximise protection for the GBR.

- Checks already undertaken on a significant fraction of the science underpinning "reef rescue" indicate that there may be serious flaws in some of the results.
- Further checks are thus advisable so that funds can be spent most efficiently and appropriately to save the GBR.
- It is estimated that it would cost around $5 million to institute replication tests on the most important scientific results.
- This represents just 1% of "reef rescue" funds.
- We propose that an Institute for Policy Science Quality Control (IPSQC) should be commissioned to undertake this task. It will be essentially a scientific auditing organisation.
- The IPSQC should be independent of the scientific organisations whose work it will check.
- The IPSQC should be run through the Auditor General's office as that organisation, although not familiar with science, understands how to ensure audits are independent.

To address the question of how to check Policy-Science more generally than the GBR, it is suggested that 5% of funding presently allocated to the Australian Research Council and National Health and Medical Research Council funds (*ca.*$2 billion per annum in total) should be spent on replication and checking of policy science.

This proposal would represent a major innovation in the systems of science QC and would further improve the reliability of science in general.

# Background

## The Science "Replication Crisis" and "Policy-Science"

Since the early beginnings of Science in the time of the ancient Greeks, the scientific method has completely revolutionized human existence and almost always for the better. Science has progressed by constant checking, replication, argument and improvement. In some areas of science, such as Newton's laws of motion, checks are effectively done billions of times every day when people fly in a plane, drive a car or walk across a bridge. Newton's laws of motion are so well tested, checked and replicated that we stake our lives on them. But some science results are not massively validated in this way and are thus not as reliable.

Here we focus on the extent to which "**Policy-Science**" is checked, tested and replicated. We define the term "**Policy-Science**" to mean all science used as the basis for making expensive or important decisions by governments to make and deliver their policies. The critical distinction between policy-science and the rest of science is the active use by government to make expensive and important decisions on behalf of the public.

The "**Replication Crisis**" is the recent revelation from a wide range of the scientific literature that there may be major systemic failing in science Quality Control (QC) (Ioannidis, 2005, 2014). Perhaps the most high-profile example comes from the biomedical sciences where in checks made on peer-reviewed science, around half of important papers are found to be wrong. Prinz et al (2011) of the German drug company Bayer found that 75% of the literature used for potential drug discovery targets is unreliable. This issue has come to some international prominence:

> *"A rule of thumb among biotechnology venture-capitalists is that half of published research cannot be replicated. Even that may be optimistic. Last year researchers at one biotech firm, Amgen, found they could reproduce just 6 of 53 "landmark" studies in cancer research*. (The Economist, 19/10/2013).

Other authors have reported the frequency of irreproducibility at around 50% (Hartshorne and Schachner, 2012; Vasilevsky et al, 2013). It has also been suggested that false or exaggerated findings in the literature are partly responsible for up to **85% of research funding resources being wasted** (Chalmers and Glasziou, 2009; Ioannidis, 2014; Macleod et al., 2014). Despite replication studies being fundamental to establishing science reliability, such studies are rarely funded, and are not generally seen as a way of advancing a scientific career (Ioannidis, 2014).

A concern over reproducibility is shared by editors of major journals. Marcia Angell, a former editor of the New England Journal of Medicine, stated

> "*It is simply no longer possible to believe much of the clinical research that is published, or to rely on the judgment of trusted physicians or authoritative medical guidelines. I take no pleasure in this conclusion, which I reached slowly and reluctantly over my two decades as an editor of The New England Journal of Medicine*." (Angell, 2009).

The editor of *The Lancet* stated that

> *"The case against science is straightforward: much of the scientific literature, perhaps half, may simply be untrue. Afflicted by studies with small sample sizes, tiny effects, invalid exploratory analyses, and flagrant conflicts of interest, together with*

*an obsession for pursuing fashionable trends of dubious importance, science has taken a turn towards darkness."* (Horton, 2015).

The financial costs of irreproducible biomedical research are significant. Freedman et al (2015) estimated that the cumulative prevalence of irreproducible preclinical research exceeds 50%, and, in the United States alone, results in approximately US$28 billion per annum spent on research that is not reproducible.

In the light of the replication crisis it would be prudent to examine whether similar problems occur in the environmental sciences, and in particular for the policy science for the Great Barrier Reef. Indeed, a call for "organized skepticism" to improve the reliability of the environmental marine sciences has already been made by Duarte et al. (2015) and Browman (2016). In particular, Duarte et al. (2015) argue that some of the major threats to ocean ecosystems may not be as severe as is portrayed in some scientific accounts, and that

> *"the scientific community concerned with problems in the marine ecosystem [should] undertake a rigorous and systematic audit of ocean calamities, with the aim of assessing their generality, severity, and immediacy. Such an audit of ocean calamities would involve a large contingent of scientists coordinated by a global program set to assess ocean health."*

## What Quality Control Processes are used in Science?

Peer Review in Science

Peer review is the primary QC procedure used for most science results including for policy-science. The public often think that peer review involves a long and thorough process where perhaps a dozen scientists work for months to check another scientist's results. But in reality, peer review may only be a quick read of the work for a few hours by a couple of anonymous referees selected by a journal editor. This is clearly not sufficient QC upon which to base decisions that may be worth millions of dollars. The cursory nature of peer review is at least partly responsible for the Replication Crisis and some eminent scientists are very scathing about its effectiveness. For example, an editor of The Lancet, Horton (2000), commented that

> *"the system of peer review is biased, unjust, unaccountable, incomplete, easily fixed, often insulting, usually ignorant, occasionally foolish, and frequently wrong'"*

Whilst this may read as an over-dramatic appraisal, the general point is clear - it is hardly a credible system upon which to base important public policy decisions worth millions or billions of dollars

"Industry Science" QC processes.

Industry and private enterprise rarely use solely peer review as its QC process. The possibility of considerable financial losses arising from an ill-informed decision is likely to drive a far more rigorous analysis to check the data and replicate results. Consider a pharmaceuticals company wishing to take a promising laboratory discovery to produce a new prescription drug for the market. On average, this costs 2.5 billion US dollars (figures for 2014) and takes a decade or more (DiMasi et al., 2016), so drug companies take great care at the beginning of a programme to make sure the initial information upon which they are basing the investment is sound. It is not accepted that a peer-reviewed journal article is adequate, partly because when checks are made, the original work is found to be wrong at least half the time (Prinz et al, 2011), and so by identifying the errors early, any waste of resources is minimised.

Engineering is far more advanced than most sciences in developing rigorous QC processes – indeed there are international standards that describe such processes in great detail. One reason that Engineering is generally more reliable than the sciences is that when an Engineer makes a mistake, there is the possibility of loss of life or massive financial loss. Consider the hypothetical case of an engineer who claimed to have invented an advanced metal alloy that would improve aircraft jet engine performance. An aircraft engine manufacturer would not regard a peer reviewed publication on the metal alloy as sufficient evidence to immediately use the metal alloy in passenger aircraft. The consequences of failure are disastrous and life threatening. Instead the engine manufacturer would subject the metal alloy to very stringent quality tests, under the guidance of strict regulations, and after many years it might ultimately be certified to use the metal for general use. This process is far more stringent and reliable than peer review.

The differences between the governments innocent trust in peer review and the far more rigorous systems of QC used by industry and engineering is remarkable. Given the level of government spending and significance of associated policy decisions, we believe that governments need to subject policy-science to greater scrutiny than the standard peer-review processes.

## Great Barrier Reef Policy-Science

Both the Queensland and Australian Governments have already spent considerable sums on the Great Barrier Reef region, including AUD $375 million between 2008 to 2013 (Reef Water Quality Protection Plan Secretariat, 2013) and are expected to spend a further AUD $575 million in water quality initiatives between 2015 and 2020 (Great Barrier Reef Water Science Taskforce, 2015; Great Barrier Reef Marine Park Authority & Queensland Government (2015); Kroon et al., 2016). More recently the Commonwealth government has announced another $500 million to be spent on reef rescue. These costs do not include those borne by industry in meeting environmental legislation or the opportunity costs of preventing some forms of development in GBR river catchments or at the coast. Such costs are difficult to estimate, but by itself, the government expenditure of AUD $1 billion or more warrants rigorous scrutiny of the science well beyond peer review.

## Significant concerns over some GBR policy-science

We have examined some of the most highly cited policy-science papers (Larcombe and Ridd, 2018) which have asserted damage to the GBR. Some of these policy-science papers make very significant claims about the health and the future of the GBR system, including dire predictions of the imminent demise of the GBR. These papers make a wide suite of conclusions directly relevant to policy, which our analysis (Larcombe and Ridd, 2018) indicates should be viewed with some doubt. These include:

(a) Riverine discharge is significantly increasing GBR water turbidity.

(b) Nutrients from agricultural runoff are largely responsible for Crown-of-Thorns starfish plagues.

(c) Pollution from agricultural runoff is affecting the species diversity of the reef.

(d) There was a 50% reduction in coral cover in the GBR from the early 1960's to 2000.

(e) There was a 14% reduction in coral growth rates between 1990 and 2005.

(f) Coral cover will fall to 5%-10% by 2022.

(g) The outer and inner GBR are 28% and 36%, respectively, down the path to ecological extinction.

Although there are a large number of papers which claim some degree of 'stress' on the GBR system, associated with increased fluvial loads, dredging, higher temperatures, lower pHs, higher chlorophyll concentrations and other parameters, the number of papers which assert to document a measurable decline of the GBR system's coral is very small. It is not suggested that all of this small number is erroneous, but in order to make reliable decisions, we must first determine what science is sound and what is not.

For the GBR, there are a range of perceived 'threats' and a limited financial capacity of governments and industry to address the problems. It is likely that some threats are far more important than others, and there should be carefully focussed expenditure on 'remediation'. However, in the light of the replication crisis the risk is considerable that the present focus of remediation efforts is misdirected.

## How to achieve rigorous technical scrutiny for GBR policy-science?

We propose that governments should establish a new independent organisation to undertake quality reviews and audits of important scientific results which underpin government spending decisions – an Institute for Policy-Science Quality Control (IPSQC). We suggest it could conduct a system of guaranteed and organised technical debate, with the aim to specifically and rigorously test for any significant deficiencies in the scientific work upon which the major public expenditure is based.

The precise mechanisms used by this organisation could take a number of different forms. However, whatever the mechanism(s) used, there must be independence, openness and transparency in all aspects. As an example, in an adversarial model, the organisation might act like a defence attorney in a court trial, challenging the scientific evidence being used to support the government decision. Depending upon the specific cases, this is likely to involve questioning scientists, commissioning attempts to replicate previous work, reanalysing data, checking experimental design, analytical methods and results, and ensuring that alternative interpretations are thoroughly considered.

Under this model the IPSQC would not adjudicate on a particular issue, but provide independent, transparent evidence, gap analyses, technical scientific counter-arguments and other advice, to support the decision-making process. The decisions themselves would, as now, be made by relevant parts of government and ministers. Whatever form the IPSQC might take, it should work in a common way on all relevant policy-science. For the case of the GBR, some of the papers that assert damage to the GBR are probably wrong and should be checked because they are being used to inform government decisions. However, there is no reason why, in the future, the situation could be reversed, and there may be scientific results which indicate that a particular perceived threat to the GBR is *not* important, which should be subject to the same level of scrutiny.

Given the national and international significance of the GBR system, a government ought not to act to formulate policy or make spending decisions on this or any other significant conclusion without making sure it is subject to appropriate rigorous technical scrutiny. As well as improving the clarity of the evidence upon which government policy and spending is made, it will improve science quality in the long run. The benefits of taking such an approach far outweigh the risks of not improving matters.

In addition to the benefits to government, business/industry will also experience flow on benefits through:

- the reduced risk of unnecessary environmental legislation being implemented and its associated compliance costs, and
- the scope for government to modify existing environmental legislation to reflect the greater system understanding that will arise through the checking and audit process.

## Funding GBR Policy Science checking

For the GBR, we estimate that it would cost around $5 million to do initial checks of many of the main scientific results upon which the "reef rescue" policies are based. This is a very small sum relative to the funds that have been allocated to save the reef such as the latest Commonwealth government allocation of $500 million. Thus only1 % of allocated funds would need to be spent on checking the underlying science. We believe that after the checking has been completed, there would very likely be a reconsideration of the priorities from threats that are less important to those where a greater improvement in reef health can be achieved. The return on the 1% investment in terms of environmental benefit is likely to be very large.

## Funding Policy Science in general

For the wider problem of checking policy science, the Australian Research Council (ARC) and the National Health and Medical Research Council (NHMRC) as the commonwealth governments primary science funding entities should become involved. The ARC and NHMRC have budgets of *ca.* $800 million and $1100 million per annum respectively and we propose that 5% of this $2 billion budget ($100 million) should be allocated to replicating and testing policy science that is being used by Australian governments. It is interesting to note that under present ARC funding rules, replication testing is probably not eligible as the ARC defines research as

> *"research is defined as the creation of new knowledge and/or the use of existing knowledge in a new and creative way so as to generate new concepts, methodologies, inventions and understandings."*

As replication and checking is not necessarily producing "new knowledge" it does not necessarily fall under the definition of fundable research.

It is thus clear that the cost of checking and replicating policy science can be done for a very small fraction of the current budget allocated to research. It should be noted that the Dutch and US equivalents of the ARC, (Royal Netherlands Academy of Arts and Sciences, and the National Science Foundation) have both recently instituted policies to do checking and replication studies.

## How can an IPSQC be established to be an efficient and independent science auditing institution

The IPSQC is effectively an auditing institution not unlike financial auditing organisations used for checking both government or corporate accounts. One of the main principles of auditing is that the auditors must be independent of the organisation to be audited. Due to its experience with independent audits, the IPSQC might be operated through the Auditor Generals department. It must not be possible for the IPSQC to be "captured" or unduly influenced by the organisations whose work it will check. In addition, the IPSQC must not become a self -serving monster that continuously advocates for its own cause and pointlessly requires more and more funding to check what has become very well-established science, while ignoring other more pressing problems.

The IPSQC would not attempt to do all the checking and replication studies itself but would commission other independent scientists or science consulting companies to do much of the work. One of its primary functions would be to identify the main papers in the scientific literature upon which a government policy is based. For the case of the GBR, this is a simple matter as there is considerable documentation of the various threats that allude to scientific papers.

We will not go into too much detail about how the IPSQC might be established to remain both independent and efficient. It would be worthwhile to consider the following points.

- How do financial auditors stay independent from accountants they are checking?
- How does the legal system guarantee a genuine argument where there is always both a defense and prosecution lawyers?
- How does the legal system guarantee no collusion between the defense and prosecution?
- Once the IPSQC is set up there would be individuals and companies with a vested interest in continuing the system People and organization would make a career out of checking science.
- The existence of an IPSQC would have the effect of making scientists more careful about drawing conclusions that may not be justified. Checks that are now not likely will become inevitable.
- It would become an honorable and noble thing to be employed to check other scientists work whereas nowadays it is bad for your career and funding for such activity is almost impossible to obtain.

# Conclusion

It has become obvious that there is a need for a considerable improvement in the QC mechanisms used to ensure reliability of policy science. In this document we outline the problems and suggest a solution. These are

- The establishment of an independent Institute for Policy Science Quality Control.
- For the GBR, 1% of 'reef rescue" funds should be spent on science checking and replication studies.

- For other potential problems with policy science 5% of Australian Research Council and National Health and Medical Research Council funds should be allocated to checking and replication studies.

The IPSQC must be established in such a way that it remains independent of the organisations whose work it will audit. The possibility of "capture" is real so considerable thought would need to go into the details of its charter and organizational structure.

## 7. References.


Angell, M (January 15, 2009). Drug Companies & Doctors: A Story of Corruption . *The New York Review of Books*.

Browman, HI (2016) Applying organised scepticism to ocean acidification research. ICES Journal of Marine Science, 73(3), 529–536. doi:10.1093/icesjms/fsw010

Chalmers I, Glasziou P. (2009) Avoidable waste in the production and reporting of research evidence. *Lancet*, 374: 86–89.

Commonwealth of Australia (2015) *Reef 2050 Long–Term Sustainability Plan*.

DiMasi JA, Grabowski HG, Hansen RA (2016) Innovation in the pharmaceutical industry: new estimates of R&D costs. *Journal of Health Economics*, 47:20-33.

Duarte, C. M., Fulweiler, R. W., Lovelock, C. E., Martinetto, P., Saunders, M. I., Pandolfi, J. M., Stefan, G., et al. 2014. Reconsidering ocean calamities. *Bioscience*, 65: 130–139.

Great Barrier Reef Water Science Taskforce (2015) Great Barrier Reef Water Science Taskforce. Full Interim Report - December 2015. Clean water for a healthy reef. pp. 67. Brisbane, Australia, Queensland Government Department of Environment and Heritage Protection.

Hartshorne JK, Schachner A (2012) Tracking replicability as a method of post-publication open evaluation. Front Comput Neurosci 6: 1–13. doi: 10.3389/fncom.2012.00001 PMID: 22291635

Horton, R., 2015. "Offline: What is medicine's 5 sigma?" *Lancet*, 385: 1380

Horton, R. (2000) Genetically modified food: consternation, confusion, and crack-up. *Med. J. Aust*. 172, 148–149.

Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLoS Medicine*, 2(8), e124. Retrieved from http://www.plosmedicine.org/article/info:doi/10.1371/journal.pmed.0020124

Ioannidis JPA (2014) How to Make More Published Research True. *PLoS Med* 11(10): doi:10.1371/journal.pmed.1001747

Kroon, FJ, Thorburn P, Scaffelke B and Whitten S (2016). Towards protecting the Great Barrier Reef from land based pollution. *Global Change Biology* 22, 1985-2002.

Larcombe, P., Ridd, P., 2018. The need for a formalised system of quality control for environmental policy-science. *Mar. Pollut. Bull*. 126, 449–461.

Macleod MR, Michie S, Roberts I, Dirnagl U, Chalmers I, et al. (2014) Biomedical research: increasing value, reducing waste. *Lancet*, 383: 101–104.

Prinz, F, Schlange, T. & Asadullah, K (2011) Believe it or not: how much can we rely on published data on potential drug targets? *Nature Reviews Drug Discovery*, 10, 712. 10.1038/nrd3439-c1 PMID: 21892149

Vasilevsky NA, Brush MH, Paddock H, Ponting L, Tripathy SJ, et al. (2013) On the reproducibility of science: unique identification of research resources in the biomedical literature. *PeerJ* 1: e148. doi: 10. 7717/peerj.148 PMID: 24032093