

Reshaping Science Structures to Better and More Coherently Serve Aotearoa

Cohesion, continuity, critical mass, data analysis, data resources,
scientific processes and incentives, and widening the funding focus.

John Maindonald

11 March 2022

Contents

Summary	2
1 Needed changes in RSI infrastructure	3
1.1 Sharing of skills across industry sector groups	3
1.2 Requirements for research organizations	4
1.3 Collaboration and continuity	5
2 Data analysis and related skills	6
3 Open data, all data, and reproducible reporting	7
3.1 The maintenance of historical data records	8
4 Scientific critique of scientific processes	9
4.1 A manifesto for change	11
5 Artificial intelligence (AI) and machine learning	12
6 Unbiased and open advice in the public arena	14
7 Features of a new national RSI infrastructure	15
Notes	15
John Maindonald – biographical details	19
References	19

Summary

1. Challenges include the removing of current obstacles to cohesion and continuity, to the building of critical mass in key skill areas, and to ensuring that research is done with attention to all relevant areas of expertise. Science has major roles in tackling New Zealand's massive environmental and social challenges, in serving current industry, and in seeding the industry of the future.
2. Data analysis skills are among those that come from technical knowhow that have been honed by wide experience and continual upskilling, and do not suddenly appear when required. The increasing use of automated forms of data analysis makes conceptual statistical issues more than ever important. These have not changed, notwithstanding huge advances in the computational machinery that is available to collate data, to do analyses, to check on analysis output, and to report results.
3. Open data initiatives have contributed greatly to advances in science, and need to be extended much more widely. Attention is needed to the maintenance of historical data records.
4. Funding, promotion, and publication structures have, in important research areas, worked to encourage getting papers into print, rather than research that is replicable and/or otherwise a genuine contribution to scientific advance. This needs to change. Funding agencies have an important role to play in insisting that, in experimental work, an insistence on independent replication becomes standard practice.
5. Mechanisms are needed that mitigate risks that private or political interest will over-ride public interest.
6. All research should be seen as part of an ongoing historical process that builds on the past, and that adds to resources of skill, data, instrumentation, and scientific understanding.
7. I regard it as important to identify projects that may assist in addressing historical Maori deprivation that has resulted from the alienation of land and resources.
8. Planning and working with hapu and iwi as kaitiaki, in research execution and planning as well as in steps that may follow, will help ensure that gains made are not lost.
9. Mechanisms are needed to ensure that funding agencies are able, and do, look closely into statistical design and analysis issues for research proposals. A plan for independent replication should, certainly in those areas of work where replication studies have demonstrated serious issues, be built in where experimental results will be used as a basis for claims that have real world implications. There should be regular reviews, across a year more, of published experimental and observational ("population-based") studies that appear in Royal Society journals.

1 Needed changes in RSI infrastructure

Whatever new Research, Science and Innovation infrastructure emerges, collaboration needs to extend well beyond that of “Science New Zealand.” Current structures divide up the public good scientific enterprise in ways that overly reflect industry sector interests when Crown Research Institutes were set up in 1992. Industry sector interests would now divide up differently, with greater changes in prospect in the next decade and beyond. The division of work by industry sector has at the same time fragmented the skill base, creating obstacles to shared learning, to transfer of skills, to cooperation, and to sharing the benefits of lessons learned and experience gained.

At Mount Albert Research Centre where I was at the time, we lost benefits of a shared library. For my area of work, we had to rely on our personal libraries.¹ It would have made sense to negotiate access to University library facilities. Also unfortunate was that what had been a common tearoom ceased to be a meeting place for scientists at the centre. There may well have been gains from the freedom that individual CRIs gained to pursue their own path. Important international links have been established. Joint appointments have worked to create channels of communication with universities.

1.1 Sharing of skills across industry sector groups

The Waikato DHB incident in May 2021 highlighted the need to ensure that, in all areas of Government and corporatized Crown entities (“quangoes”), IT system security is managed by experts with the best available levels of skill. The principle applies, albeit with less force, to database design and maintenance, and data analysis skill areas. Except where there are particular specialized requirements, it makes sense to share these across whatever new entities that emerge from the restructuring. They can then managed to provide working conditions that will attract really able staff, to ensure attention to skill maintenance, and to ensure that staff stay abreast of important technical innovations. The combined and complementary skills of multiple sharp and well-trained minds are likely to be more effective resource than those of any one individual.

The nature of my experience gives me a particular interest in expertise in the role of numerical computation, statistical data analysis, and database management. For data analysis, anyone with computer language skills is unlikely to have much difficulty in gaining familiarity with the R language that is widely used for, among other things, professional statistical analysis. Learning to use the abilities in those packages that may be required for specialist use is where the challenges lie.

¹Commentary in Kenna and Berche (2010) is pertinent, and for work in statistics, Kenna and Berche (2012).

For large parts of the work of GNS and NIWA, research gets close international scrutiny. Mathematical science skills are so central to the teamwork required to do anything credible, and there is such continuity of work, that skill maintenance and development is a built-in component of projects. Where such demands are not thus built in, careful management is required to balance the competing demands: to maintain continuity and organisational memory in the face of start-stop demands from clients (including MPI), to maintain the skill base, and to innovate and pioneer new directions.

1.2 Requirements for research organizations

A 2018 seminar on “retooling primary health care for the 21st Century,” impressed me with the detail of the scrutiny now being given to the effects of the organisation of patient care services on achieving good patient outcomes, and on costs.² A comparable level of scrutiny, albeit addressing very different issues, should be paid to the processes by which research organisations (the organizations, not the scientists who work for them) organize their work for scientific outcomes. Measures must be more incisive and focused on the public interest than standard forms of feedback from commercial or government customers, and more reliable than common measures of research quality.³

Other needed measures, with progress regularly reported, include:

- a strong focus on staff training and upskilling, especially in areas where published work on reproducibility has identified common serious deficiencies
- a requirement to place in the public domain all data, with details of what is available published on the organization’s website. Where there are confidentiality issues, a mechanism is required by which the data can be made available on an individual basis.

An over-riding concern is that all research, both in science organisations and in Government, should be seen and reported as part of an ongoing historical process. More than specific scientific results are at stake — all worthwhile research adds to an ongoing body of data, of resources, and of scientific understanding. Data, and research results, are taonga that future generations should be able to use with some reasonable confidence as a points of departure for their own research.

²www.victoria.ac.nz/events/2018/04/health-care-home-retooling-primary-health-care-for-the-21st-century

³On measures of research quality, see Kenna, Mryglod, and Berche (2017).

1.3 Collaboration and continuity

Areas of expertise for which there is a demand across several CRIs would in many cases be better shared within a dedicated agency, where a strong focus can be placed on maintaining a critical mass – for recruiting, for mentoring, for skill development, and for provision of advice. It is in any case wasteful for multiple agencies to duplicate the effort and inputs required to develop resources and skills that would better be developed jointly, with a better result.

Projects proceed in a stop-start manner will often need to rely on consultants who are brought in for the duration of the project. This increases the risk that skills and organisational learning will have dissipated when and if closely work is taken up at some later time. I judge this to be a particular issue in biosecurity. As with Covid-19 and other human pathogens, it is important to move beyond ad hoc responses to issues as they arise to preparing for new challenges that will arise in the future. See Note 1.

Biosecurity shares challenges that are common with those for human pathogens: in maintaining records, contact tracing (e.g., for farm animals), and modeling spread of disease. Connections with human health issues surely warrant careful attention, especially as in the case of animal health issues such as created by *Mycoplasma bovis*.

Features of the changing scene are new data sources, often giving very large datasets, automated “machine learning” type mechanisms for processing data, and advances in the tools available for statistical modeling. It has, unfortunately, become easier than ever to identify spurious as well as real associations, deceiving untrained human intuition in ways such as are documented in Kahneman (2011).⁴ Experts in a specialist area may, because of gaps in their understanding, make serious mistakes of judgment. Sally Clark’s wrongful 1999 conviction in the UK for the murder of her two children, overturned only after her life had been ruined by three years in prison, is an extreme example of what can go wrong.⁵

Managers need to understand these issues well enough to ensure that, as need may arise, they seek comment and/or help from suitably skilled advisers. Attention to these issues becomes more than ever important as tasks that were previously handled manually are automated.

⁴See also Smith (2014). This entertainingly written book comments on examples, from published papers and from the media, of common types of data misinterpretation.

⁵“Expert” evidence from a pediatrician who apparently had not seriously considered that “cot death” risk was likely to vary between families, would appear to have been a major contributor to the guilty verdict. Why did the legal experts involved, including the judge, not pick up on this point? See https://en.wikipedia.org/wiki/Sally_Clark

2 Data analysis and related skills

As with the delivery of medical and social services, there is a need for to structure delivery services in ways that will effectively serve the needs of client groups. Where this same structuring is carried across to skill maintenance, this fragments the skill base.

While statistical analysis training in the university sector has made important strides in the past several decades, ongoing professional training that builds on that initial training and keeps pace with technological change has not received the same attention. New opportunities to link in via the internet to meetings and workshops in Australia and elsewhere are not an adequate replacement for the seminars presented by statistical analysis experts from abroad that the former DSIR Applied Mathematics Division organized from time to time, both for its own staff and for professionals in other organizations. Statisticians working in science and Government do not link, to the extent desirable, to the wider statistical community. At least in Wellington, meetings that bring together statistical specialists with an application area interest are now very occasional events.

Close liaison with the scientists with whom biometricians and other mathematical problem solvers work is important, but so also is the stimulus that comes from exchanging insights with and learning from those who work in closely related specialist areas. Work across multiple application areas, and contact with the work of others in the larger group, provides a valuable vantage point from which to see the scientific enterprise as a unified whole. This has especial value in early career. Methods used in one area will often be useful in another. Much methodology that was developed for use in one application area has, in the modern history of theoretical statistics, been re-invented under another name for use in another area.

There is a strong continuity of interest with medical biostatistics. Scientific work in other areas (including, ironically, pre-clinical laboratory studies) can usefully learn from the approaches that the Cochrane Collaboration follows in its studies, from the important role of meta-analyses, and from genuinely evidence-based approaches more generally.⁶

Approaches to the assessment and management of risk that have been pioneered in medicine have come strongly into public view in the course of the Covid-19 pandemic. These, suitably adapted, find wide application across all areas of public life – parole board decisions, earthquake and tsunami risk, fire risk, the risk of miscarriage of justice, public health risk from pollution, risk of vehicle or aircraft accident – the list is endless.

⁶Links for NZ and international Cochrane websites are <http://nz.cochrane.org/> and <http://www.cochrane.org/>

Some science funding needs to go into public education, cooperating closely with the Ministry of Health, with Cochrane NZ, and with the new public media entity. The Covid-19 epidemic, and the protests at Parliament have highlighted a growing undergrowth of misinformation that has created serious issues, for trust in established science more generally as well as for public health. The tradeoff of costs and benefits that arises in any public health contexts need to be widely understood. Critical thinking, with a strong focus on the statistical issues, needs to become a strong focus in teacher training and in the school curriculum.

3 Open data, all data, and reproducible reporting

Since the latter part of the last century. there has been a huge expansion, in many research areas, in the databases that are available to researchers, making possible advances that would not otherwise be possible. Areas where this is particularly obvious include climate science, earthquake science, geology more generally, and molecular biology. In molecular biology, they have had a pivotal role in the technology that allowed a much more effective response to the Covid-19 pandemic than would have been possible two decades earlier. The 2022 annual Nucleic Acids Research database issue (Rigden and Fernández 2022) notes 1645 database entries in its online collection. Seven of the 89 new databases listed related to Covid-19 and the Covid-19 virus.

Areas where the gains that stand to be made are less obvious have not availed themselves of the new opportunities. Making data publicly available places a strong discipline on those who create the data, and helps ensure appropriate documentation. It opens it up to wider use, where PhD and post-doc students can expose it to methodology that they may be in the process of developing, and can look for features that may have been missed or misrepresented in the published analysis, or that earlier modeling software was unable to handle well. It increases the chances that data will be preserved in a form that can be used by posterity. See further, Note 2.

Data that is in the public domain will from time to time attract the attention of “citizen science” data analysts. Depending on the nature of the data and on the background information needed to use it effectively, such data will from time to time find its way into the hands of analysts who have the skills needed to do a really effective analysis job. Another model that can work well in some contexts comes from the Kaggle⁷ organisation’s success in making predictive modeling, with data provided publicly, a commercial enterprise.

Issues that relate to the use of commercial databases, with claims of commer-

⁷<https://www.kaggle.com/>

cial sensitivity used as reason for not making data available for any external check, came into strong focus when a May 2020 Lancet article appeared where results were stated to be based on data from 700 hospitals across six continents that been provided by the healthcare company Surgispere. The article quickly attracted a letter from 100 scientists worldwide that raised a number of methodological and data integrity, and ethical concerns. The article was retracted two weeks after publication. Note 3 has further details.

Where possible, all relevant data that bear on an issue under investigation should be brought together, from international as well as from New Zealand sources. In clinical medicine, the use of a “meta-analysis” to bring together evidence from multiple studies has become common practice. These can be much more effective if the data on which the studies were based is available.

Reproducible reporting, so that all code used to handle analyses and produce figures and graphs is available to allow others to check and/or repeat or update what has been done, although a harder ask, should be mandated, albeit with access limited if there are over-riding confidentially concerns. There are large benefits – updates to reports or papers are straightforward, with minimal risk that new errors will be introduced. It makes it more likely that errors or omissions will be identified. See Note 4

3.1 The maintenance of historical data records

An obituary for glaciologist Trevor Chinn (d. 20 December 2018) drew attention to his meticulous work in recording and keeping together data on Southern Alpine glaciers that would otherwise have been lost in the course of successive public sector restructures.⁸ In other cases, without a Trevor Chinn to maintain the data intact, important data has been lost in the restructuring and downsizing of government agencies has led to serious losses of historical data, seriously compromising current and future work for which it would have added important insight. Paper resources have often been “recycled.” It is ironic that, while the National library has responsibility for “collecting, preserving, and protecting documents” relating to New Zealand, and for making them available, there is no body that has any comparable responsibility for maintaining collections of historical scientific data. Note also the work that the Department of Statistics does in making its data publicly available.

- In three projects in which I was involved after returning to New Zealand in 2015, data that I had worked on in the 1990s and that would have been helpful both for checking new analysis methodology and of interest in its own right, was not available. While this was at least partly a

⁸<https://www.stuff.co.nz/environment/109932454/life-story-trevor-chinn-the-man-who-saved-glaciology>

result of circumstances over which the nascent CRIs had no control, there is no such excuse for data that has been collected more recently.

- At least in some places, it is still left to scientists to maintain their own data in what will often be messy spreadsheets, with no proper versioning, and no check that documentation is adequate. See further, Note 5.

In GNS and in NIWA, there has been good work in setting up databases, with much of the data publicly available. Why does this not happen more widely in science? It has to be emphasized that databases have to be maintained in ways that ensure continuing access as technology and the demands on them change.

The needed resource(s) would be best shared across science agencies, giving at least limited protection against the losses of data and historical records that have in the past accompanied restructuring of public sector agencies. Genomics for Aotearoa New Zealand (GFANZ)⁹, set up to facilitate the sharing of genomics data between New Zealand researchers, may be a useful model for what is needed more widely.

Biological scientists are among those who will commonly not be comfortable moving data from the Excel spreadsheets to which they are accustomed to the style of database needed for long-term storage (though this is changing with a new generation of graduates.) Specialists may be required who can take over the work, or at least help.

4 Scientific critique of scientific processes

Reports internationally on the extent to which the majority (in some cases, the great majority) of published experimental studies have proved irreproducible make a clear case for paying much better attention than in the recent past to the dependability of published work. Published work that is not replicable wastes the time and resources of those who try to build on it. It is then a serious concern that in studies of published results from laboratory experiments have shown reproducibility rates that have commonly been at best around 40%, and at worst as low as 12%. Areas covered include pre-clinical medicine, psychology, and laboratory economics. See Note 6. A recently concluded study that attempted to replicate 193 experiments from the 53 “most impactful” cancer biology studies from 2010-2012 was able to replicate only 50 experiments, from just 23 of the 53 papers. Note 7 gives summary details.

The inevitable gap between results from work that has sufficient credibility to warrant further investigation, and well-established results such as those

⁹<https://genomics.nz/>

that underpin approvals for the use of vaccines, is not an adequate excuse for very low rates of reproducibility of published work.

Thus, in laboratory studies, refereeing processes have in the recent past done little to ensure scientific credibility. Studies that “simply” try to reproduce the results of others have, in many areas, not been considered for publication. This needs to change. Insistence on independent replication of laboratory results should be standard practice. P-values or other statistical measures are important as adjuncts to independent replication, but are not a substitute. Replication places a focus back on all aspects of the experimental process – experimental design, experimental procedure, and the quality of statistical analysis – in ways that no other mechanism can.

In a paper entitled “Cargo-cult statistics and scientific crisis,” Stark and Saltelli (2018), comment:

Statistics was developed to root out error, appraise evidence, quantify uncertainty, and generally to keep us from fooling ourselves. Increasingly often, it is used instead to aid and abet weak science.

There is no lack of work that melds effective use of statistical methodology with strong science. That melding should be the standard for all areas of statistical application. The challenge is to ensure that statistical analysis gives insightful and defensible results, however the contexts for that challenge may change and widen.

The poor quality of experimental design and of statistical analysis in much published work is addressed in scathing terms in Collins and Tabak (2014):

Factors include poor training of researchers in experimental design; increased emphasis on making provocative statements rather than presenting technical details; and publications that do not report basic elements of experimental design. Crucial experimental design elements that are all too frequently ignored include blinding, randomization, replication, sample-size calculation and the effect of sex differences. Exacerbating this situation are the policies and attitudes of funding agencies, academic centres and scientific publishers. . . .

[Following a distinguished career in medicine, Francis Collins became Director of the US National Institutes of Health in 2009, a post from he retired at the end of 2021.]

C. G. Begley (2013) commenting on the report in C. Begley and Ellis (2012) that Amgen scientists had been able to replicate 6 only of 53 ‘landmark’ cancer studies, identified very similar issues (“Six red flags”), commenting

What is also remarkable is that many of these flaws were identified and expunged from clinical studies decades ago. In such studies it is now the gold standard to blind investigators, include concurrent controls, rigorously apply statistical tests and analyse all patients — we cannot exclude patients because we do not like their outcomes.

There should be regular reviews of published work and associated online documentation, focusing on issues such as those identified in Collins and Tabak (2014) and C. G. Begley (2013). I am not aware of any overview of the statistical content of published work in New Zealand biological journals, comparable to Maindonald and Cox (1984), that has appeared since that paper. Reviews of this general type do appear in the international medical literature from time to time. See, e.g., Parsons et al. (2012), and papers cited there.

As Collins and Tabak (2014) argue, funding agencies have an important role to play in making scientific processes more scientifically credible. Greater use of expert statistical would help, both in study design and in statistical analysis.

4.1 A manifesto for change

Munafò et al. (2017) is a manifesto for change.¹⁰ Proposals are wide-ranging in their implications. There should be pre-registration of study design, primary outcome, and analysis plan. Methodological training and support should be a strong focus. Team collaboration should be encouraged. Reporting guidelines have an obvious role, but will not on their own be enough to address reporting biases. Review processes can and should be extended to include public forms of both pre- and post-publication evaluation and review. Reward structures that look broadly at researchers' output and focus less on apparent novelty would change researcher behaviour in positive ways. See also the wide-ranging critique in Ritchie (2020).

There is, in at least some areas of public health research, an over-confidence in what can be done using regression approaches on data from observational (“population based”) studies. Claims that are based on such modeling require rigorous critique. The use of new tools for collecting data, in public health and in other areas of society and government, combined with the use of machine learning tools to automate attempts to extract meaning from the data, will open new opportunities for over-interpretation and/or misinterpretation. See Note 8.

It is ironic that both climate science and the science that underpins the

¹⁰See also Higginson and Munafò (2016) on the effects of reward structures.

safety and effectiveness of the vaccines – areas where widely drawn scrutiny and critique have ensured that standards are high – have been prime targets of sustained attempts to undermine credibility. In climate change especially, work almost inevitably requires co-operation between individuals with different areas of technical expertise who can be expected to look carefully over each other’s work. The safety and effectiveness of Covid-19 vaccines have had extraordinarily high levels of testing and checking, both in clinical trials and in day to day use.

Cases where drug manufacturers has been able to manipulate their way around US Food and Drug Administration requirements destroy trust in pharmaceutical approval processes more generally. There was a huge increase in drug overdose deaths in the United States from 6.1 to 21.6 per 100,000 between 1999 and 2019.¹¹ A major factor was the increased use of prescription opioids. Purdue Pharma stands out for its aggressive marketing of oxycodone, sold under the brand name OxyContin, arguing that concerns over addiction and other dangers from the drugs were overblown.(Kolodny 2020)

5 Artificial intelligence (AI) and machine learning

The term “artificial intelligence,” used as a catch-all for a variety of types and applications of algorithmically based automation, is unfortunate. It suffers from the same potential as human minds to make mistakes. These reflect, however indirectly, flaws that result from faulty coding and/or issues with the data used. Fry (2018) is a fascinating overview, accessible to the lay reader, with a huge range of what often read like well-told detective stories to illuminate the exposition. Fry comments that:

. . . the hype over AI is a distraction from much more pressing concerns and — I think — much more interesting stories

Currently in view are many different forms of “narrow AI.” The term “machine learning” makes good sense for systems that allow machines to act autonomously – here note robotics systems, and automated guidance systems such as are in use for aircraft and for self-driving cars. These take and use feedback directly from the environment. System failures, albeit with potentially catastrophic results, are directly obvious. As an example of the very serious consequences that may follow, see Note 9.

Contrast such autonomous systems with systems that, relying on data supplied to them, or on data that have been extracted from administrative records, are designed to assist or direct decision-making.¹² Great care is

¹¹<https://www.cdc.gov/nchs/data/databriefs/db394-tables-508.pdf#page=1>

¹²This use of machine learning algorithms is a form of “regression,” with the same scope for mistakes in use and in interpretation of output. Where the interest is accurate

needed to ensure that automated systems do not offer much increased scope for missing or ignoring issues with the available data, for allowing faulty analysis to go unrecognized, and for misinterpretation of analysis results. Without access to the data, and without clear explanation of the criteria used, there is no way to expose unfairness that is built into the data used, or manipulation of the data, or mistakes. Failures of human intuition, such as are documented in Kahneman (2011), can readily find their way into automated systems.

In the colourful language of the title of O’Neil (2016), automated systems readily become “Weapons of math destruction.” O’Neil documents issues for systems that control the deployment of police resources, or that determine hiring and firing decisions, or that may be used to drive public health decision-making.

The article Lazer et al. (2014) is an interesting commentary on Google’s attempts, over 2008 - 2013, to use their own algorithms with data collected from the web to predict flu outbreaks. The authors argue that while a statistically informed use of data from the web can usefully supplement other sources of insight, it cannot be an effective replacement for the use of data sources that more directly indicate flu incidence.

Most ‘AI’ practitioners have to date come from a training in computer science, with limited exposure to statistical issues that arise for the collection, analysis and use of data to extract meaningful information, e.g., for setting policy. It may be hoped that the new “data science” courses that are becoming common in statistics and computer science departments will incorporate substantial practically oriented statistical theory and analysis components. See Note 10.

In the attempt to extract meaning from available data, subtleties are readily overlooked. As argued at length in Barrowman (2018), in an article that makes a number of important points, no data is ever totally “raw.” The processes that generated it, and the wider context of scientific understanding, have large implications for the conclusions that it can be used to draw.

In public policy, the available data is, much more commonly than officials recognize, not an adequate substitute for the data that is needed for sound judgment. This point has relevance to measures used to assess the worth of scientific research.¹³

In assessing the effectiveness of cancer treatment, it is clearly simplistic to assume a direct link from number of apparent cancers found to lives saved. The increased chance of detecting a life-threatening cancer must be set against

prediction, it may be termed “predictive modeling.”

¹³Kenna, Mryglod, and Berche (2017)

the increased risk of serious side effects or hastened death from detecting and treating an apparently cancerous tumor that will never cause serious harm.¹⁴ As tests are developed that have an increasing ability to detect apparent “cancers,” these considerations can only increase in importance. It becomes increasingly important to educate the public to understand the trade-offs.

Some patient organisations, and some medical specialists, have seemed unwilling to argue the case in terms of balance of risk. In addition to the references just noted, see Note 11. The issues are important, both for effective use of public resources, and for avoiding tests and treatments that are on balance likely to harm patients.

6 Unbiased and open advice in the public arena

It is strongly in the public interest that scientists have reasonable freedom for responsible expression of their minds on issues of public concern. In an informal 2015 survey, 151 CRI scientists (out of 384 who responded) answered yes to the question “Have you ever been prevented from making a public comment on a controversial issue by your management’s policy, or by fear of losing research funding?” Hon Joyce’s response was an evasion, in effect arguing that as this was not a scientific survey of all CRI scientists (to this extent, true), its evidence of large concern could be ignored. Equally disturbing was the reaction of the NIWA management, suggesting a determination to brush the concerns raised under the carpet.¹⁵

A situation where commercial organisations can use the threat of loss of commercial contracts to prevent public comment from those who are best qualified to give it brings serious risk to the body politic. It need not be a contract with the individual expert involved, merely one involving the individual’s organisation.

This issue is crucially important in areas where current manufacturing practices create huge environmental, public health, and other such concerns. Consider, among other concerns: fossil fuel prospecting and use, environmental issues, use and disposal of plastics, and processed food manufacture.

Issues of this type, as they relate to the UK government’s handling of the Covid-19 epidemic, are documented in astonishingly forthright comments in Abbasi (2020).¹⁶

¹⁴See, e.g., Glauser (2018), commenting on Welch et al. (2018).

¹⁵<https://sciblogs.co.nz/infectious-thoughts/2015/08/28/niwa-in-astonishing-attack-on-scientist-association/>

¹⁶<https://www.bmj.com/content/371/bmj.m4425>

7 Features of a new national RSI infrastructure

Whatever is done, it is important that the several years of chaos that preceded and followed the breakup of DSIR in 1992 are avoided. If some version of the present CRI structure is retained, the RSIs that result should operate within a more cohesive and cooperative organizational structure than Science New Zealand provides. Data analysis, database design and deployment, information technology and IT security, are areas where there should be wide sharing of skills between the new entities. Links into relevant University departments would have the potential to benefit both parties.

Te Pūnaha Matatini rose to the modeling challenge of Covid-19 very effectively – notwithstanding the detailed criticisms that might be made of the models used. Its work might be expanded, or another body set up, with a brief to work with funding agencies to examine critically how research proposals match up against standards such as those set out in Collins and Tabak (2014) and [munaf0EtAl_2017](#). Its brief would extend to reviews of published studies and of reports arising from work undertaken by New Zealand agencies, with a role also in ensuring that scientists in these have access to high quality advisory services.

Funding needs to put a high priority on continuity of work, especially where environmental and biosecurity issues are concerned. Attention to the interests of hapu and iwi as kaitiaki, in research execution and planning as well as in steps that may follow, will help ensure that gains made are not lost.

Funding agencies have an important role to play in insisting that, in experimental work, that independent replication becomes standard practice. Standards for the use of observational (“population based”) data to demonstrate claimed causative effects need to be tightened. There should be regular reviews such as [Maindonald and Cox \(1984\)](#), of published experimental and population-based studies that appear in Royal Society journals.

Notes

1. A third planned stage of the fruit biosecurity project “Developing pre-approved and standardised quarantine treatments for fruit flies” in which I was involved, and that would have extracted much of the potential value, was not funded by MPI. Important aspects of the project might usefully be copied elsewhere. It highlighted a number of important questions and issues — in control of experimental conditions, in design of experiments, in the match between laboratory experimentation and commercial conditions, in statistical analysis methodology, and in the uses and limitations of analysis results. It took me some time to get up to date after a gap of 15 years from my

earlier involvement in such work, and to get on top of advances in modeling and computing resources that had occurred over that time. While it is fair that importers should bear a fair share of the cost of projects that support their business, this would be better recovered as a levy, allowing a continuity of research that would both benefit importers and better serve the public interest in maintaining biosecurity.

2. The plant biosecurity project mentioned in Note 1 brought together a large body of data from US, Australian, and New Zealand sources, creating for use in this project a database such as ought to be openly available to researchers on a permanent basis. Such work, and even more importantly biosecurity more generally, requires a whole-of-system approach in order to be extract maximum value.
3. Among other concerns raised, there had been inadequate adjustment for disease severity, temporal effects, site effects, and dose used. Data sources could not be verified, some quoted summary statistics showed remarkably little variation, and stated daily doses of hydroxychloroquine were 100 mg higher than FDA recommendations. In a YouTube video on the Lancet study, Desai commented, with astonishing bravado: “The real question here is, with data like this, do we even need a randomised control trial?”
4. On effective tools for reproducible reporting, journals that have made reproducible research initiatives, and links to other relevant sites, see the link in the footnote.¹⁷ These approaches, have made inroads in some parts of the research community, The startup cost in time and training is soon outweighed by benefits. It would be facilitated by cooperation across research organizations in the provision of training and support.
5. For the long-term maintenance of data records, as well as for immediate recording and use of data, the widespread reliance on Excel spreadsheets is a major source of error and confusion, in ways that increase with time from the initial recording of data. Spreadsheets lack mechanisms for integrating documentation with spreadsheet contents, for logging changes, and for checking data integrity. The mixing of data values that are computed using formulae with actual data adds to the complexities involved in checking and logging changes. An effective strategy can be to use spreadsheets for initial data entry only, with technical staff then transferring data to a more permanent database prior to analysis, at the same time ensuring attention to documentation.

¹⁷<https://reproducibleresearch.net/links/>

6. See, e.g., C. Begley and Ellis (2012), C. G. Begley (2013), Camerer et al. (2016), and Economist (2013). Issues of the type noted continue to get attention in Nature, in Science, and in other scientific journals that have a broad scope. Psychology appears to be ahead of other areas of science in addressing the issues raised. Results from the ‘Many Labs’ project (Klein and others 2014), where researchers from 36 different research groups set out to reproduce one or more of 13 “classical” studies, were more encouraging. Eleven studies were successfully reproduced (one weakly), with two failing.

7. A \$1.3 million grant from the Laura and John Arnold Foundation funded an exercise in replicating the 53 “most impactful” cancer biology studies from 2010-2012. In the end, 50 experiments from 23 papers were repeated.^[See the web link for T. Errington (2021), and the papers T. M. Errington et al. (2021), and Rodgers and Collings (2021)}.] In 92% of completed experiments, replication effect sizes were smaller than the original, with the median effect size 85% smaller than reported for the original experiment. Barriers to repeating experiments included shortcomings in documentation of the methodology; failures of transparency in original findings and protocols; failures to share original data, reagents, and other materials; and methodological challenges encountered during the execution of the replication experiments. The challenge to established practices has generated controversy in the research community, and highlighted questions on what constitutes replication.¹⁸

8. A study (Wernham et al. 2016) that did little more than suggest that there may be an issue that warrants further investigation was the basis for exaggerated claims in a 2016 New Zealand Listener article.¹⁹ The research did not, as claimed, “lob a grenade into the historically war-torn territory of New Zealand’s maternity care.” Even less did its results warrant the melodramatic “Alarming maternity research” and “Revolution gone wrong” that appeared on the Listener’s front cover. There are analysis tools that the authors of the study could and should have used to shed light on the likely effectiveness of the covariate adjustments. The analysis used, as a measure of social deprivation, the 2006 predecessor of the NZDep2013 index that applies to meshblocks of of around 60–110 people, not to individuals.²⁰ For

¹⁸Nosek and Errington (2020)

¹⁹Chisholm (2016): “Birth Control”

²⁰<https://www.health.govt.nz/publication/nzdep2013-index-deprivation>

papers of this type, that may be thought to have have public health or other public policy implications, it would be in the public interest to include expert commentary along with the published paper.

9. The 737 Max 8 fiasco, where crashes in October 2018 and March 2019 led in all to the loss of 338 lives, with the plane then grounded from March 2019 to November 2020. The system did not recognize failures in the sensors that provided the data needed for operation, and had inadequate provision for transfer to manual control. As with all such ‘AI’ systems, coding and testing and deployment were the work of human intelligences.²¹
10. It was disappointing that the report AI Forum New Zealand (2018) gave no direct attention to issues of statistical use and interpretation of output from machine learning algorithms, or to other concerns such as those that get limited attention in Gluckman (2017), and wide-ranging attention in Fry (2018) and O’Neil (2016). Fortunately, downsides of uncritical use of AI systems have entered much more into public discussion since the forum’s report appeared. Within the broad sweep of what the AI Forum report terms AI, the nature and seriousness of these concerns vary widely.
11. For some groups and some medical conditions, the available evidence indicates that general population screening risks starting a process that is likely on balance to do more harm than good. Some patient organizations, and some specialists, have seemed immune to such evidence. Several years ago, the Prostate Cancer Foundation website²² made misleading use of the claimed “tragic case of Graeme Pollard” to encourage men to ask for screening. This has now been removed, and a link provided to advice²³ that includes the needed warning that harms have to be weighed against benefits. Fact boxes such as on the website of the Harding Center for Risk Literacy (thus, see Fact Box on Prostate Cancer) can be an effective way to convey summary results from a careful review of the literature.²⁴ On unscientific attitudes among urology specialists interviewed in a US study, see Levitin (2015), pages 240 - 248.

²¹Boeing faced fraud charges, and paid more than \$2.5 billion in penalties and compensation. Serious failures of regulatory oversight were identified. Issues arose from the use of an add-on software system to adapt, for use with a larger engine that was placed higher up and further forward, software that had been designed and tested for earlier 737 models.

²²<https://prostate.org.nz/>

²³<https://screeningforprostatecancer.org/>

²⁴<https://www.harding-center.mpg.de/en/box/magazin1/9433-fact-boxes>

John Maindonald – biographical details²⁵

Following a first in Mathematics at what was then Auckland University College, and several teaching and lecturing positions, John Maindonald worked with other researchers, for the major part of his career, as a quantitative problem solver. He has held positions at Victoria University of Wellington, in DSIR, in HortResearch, and at The Australian National University (ANU).

Between 1983 and 1996, and occasionally after moving to Australia in 1996, he reviewed the statistical content of numerous papers that appeared in DSIR (later, Royal Society) journals, notably the *New Zealand Journal of Agricultural Research* and the *New Zealand Journal of Crop and Horticultural Research*.

The move to Australia opened up new and interesting vistas – work in the Centre for Clinical Epidemiology and Biostatistics at the University of Newcastle for 20 months, the work in the ANU Statistical Consulting Unit that gave contacts widely over the university. He joined the newly formed ANU Centre for Bioinformation Science in 2001. Following formal retirement in 2005, he was until 2021 a visiting fellow in the ANU Mathematical Sciences Institute. Between 2003 and 2015, he fronted a total of 35 short courses (most one week or less) that demonstrated the use of the open source R system for a wide range of data analysis and related purposes. These were conducted at the request of, or under the auspices of, a variety of Australian and other academic, research, and government organisations.

He is the author of a book on Statistical Computation, and the senior author of “Data Analysis and Graphics Using R — An Example-Based Approach” (Maindonald and Braun 2010). This has sold more than 11,000 copies over the three editions. A new text, derivative from the fourth edition, is in the late stages of preparation.

Upon returning to New Zealand in 2015 he was persuaded to become involved in three projects with the Plant and Food CRI, one of them the Plant Biosecurity project “Developing pre-approved and standardised quarantine treatments for fruit flies.”

See http://www.statsresearch.co.nz/john_maindonald.htm for a more detailed curriculum vitae.

References

Abbasi, Kamran. 2020. “Covid-19: Politicisation, ‘corruption,’ and Suppression of Science.” *Bmj*. British Medical Journal Publishing Group.

²⁵A more complete CV can be found at http://www.statsresearch.co.nz/john_maindonald.htm

- <https://www.bmj.com/content/371/bmj.m4425>.
- AI Forum New Zealand. 2018. *Artificial Intelligence: Shaping a Future New Zealand*. <https://aiforum.org.nz/2018/05/02/ai-forums-research-report-artificial-intelligence-shaping-a-future-new-zealand/>.
- Barrowman, Nick. 2018. “Why Data Is Never Raw. On the Seductive Myth of Information Free of Human Judgment.” *The New Atlantis, Summer/Fall 2018*, 129–35. <https://www.thenewatlantis.com/publications/why-data-is-never-raw>.
- Begley, C., and L. M. Ellis. 2012. “Drug Development: Raise Standards for Preclinical Cancer Research.” *Nature* 483 (7391): 531–33.
- Begley, C. G. 2013. “Reproducibility: Six Red Flags for Suspect Work.” *Nature* 497 (7450): 433–34.
- Camerer, Colin F., Anna Dreber, Eskil Forsell, Teck-Hua Ho, Jürgen Huber, Magnus Johannesson, Michael Kirchler, et al. 2016. “Evaluating Replicability of Laboratory Experiments in Economics.” *Science* 351 (6280): 1433–36.
- Chisholm, Donna. 2016. “Birth Control.” *The NZ Listener, October 8 - 14*, 18–24.
- Collins, Francis S., and Lawrence A. Tabak. 2014. “Policy: NIH Plans to Enhance Reproducibility.” *Nature* 505 (7485): 612–13.
- Economist. 2013. “Unreliable Research. Trouble at the Lab.” *Economist*.
- Errington, T. 2021. “Replication Study Results.” <https://osf.io/e81xl/wiki/home/>.
- Errington, T. M., Maya Mathur, Courtney K Soderberg, Alexandria Denis, Nicole Perfito, Elizabeth Iorns, and Brian A Nosek. 2021. “Investigating the Replicability of Preclinical Cancer Biology.” *Elife* 10: e71601.
- Fry, Hannah. 2018. *Hello World: Being Human in the Age of Algorithms*. W. W. Norton & Company. <http://www.hannahfry.co.uk/>.
- Glauser, Wendy. 2018. “Look Too Close and We’re All Sick.” *New Scientist* 237 (3172): 44–45. [https://doi.org/10.1016/s0262-4079\(18\)30619-5](https://doi.org/10.1016/s0262-4079(18)30619-5).
- Gluckman, Peter. 2017. “Enhancing Evidence - Informed Policy Making.” *Pmcsa.org.nz*. <http://www.pmcsa.org.nz/wp-content/uploads/17-07-07-Enhancing-evidence-informed-policy-making.pdf>.
- Higginson, Andrew D., and Marcus R. Munafò. 2016. “Current Incentives for Scientists Lead to Underpowered Studies with Erroneous Conclusions.” *PLOS Biology* 14 (11): e2000995. <https://doi.org/10.1371/journal.pbio.2000995>.

- Kahneman, Daniel. 2011. *Thinking, Fast and Slow*. 1st ed. Penguin Books.
- Kenna, R., and B. Berche. 2010. “Critical Mass and the Dependency of Research Quality on Group Size.” *Scientometrics* 86 (2): 527–40. <https://doi.org/10.1007/s11192-010-0282-9>.
- . 2012. “Statistics of Statisticians: Critical Masses for Research Groups.” *Significance* 9 (6): 22–25. <https://doi.org/10.1111/j.1740-9713.2012.00617.x>.
- Kenna, R., O. Mryglod, and B. Berche. 2017. “A Scientists’ View of Scientometrics: Not Everything That Counts Can Be Counted.” *Condensed Matter Physics* 20 (1): 13803. <https://doi.org/10.5488/cmp.20.13803>.
- Klein, Richard A., and others. 2014. “Investigating Variation in Replicability.” *Social Psychology* 45 (3): 142–52.
- Kolodny, Andrew. 2020. “How FDA Failures Contributed to the Opioid Crisis.” *AMA Journal of Ethics* 22 (8): 743–50. <https://pubmed.ncbi.nlm.nih.gov/32880367/>.
- Lazer, D., R. Kennedy, G. King, and A. Vespignani. 2014. “The Parable of Google Flu: Traps in Big Data Analysis.” *Science* 343 (6176): 1203–5.
- Levitin, Daniel J. 2015. *The Organized Mind*. 1st ed. Penguin Books.
- Maindonald, J H, and W J Braun. 2010. *Data Analysis and Graphics Using R – an Example-Based Approach*. 3rd ed. Cambridge University Press.
- Maindonald, J H, and N R Cox. 1984. “Use of Statistical Evidence in Some Recent Issues of DSIR Agricultural Journals.” *New Zealand Journal of Agricultural Research* 27: 597–610.
- Munafò, Marcus R., Brian A. Nosek, Dorothy V. M. Bishop, Katherine S. Button, Christopher D. Chambers, Nathalie Percie du Sert, Uri Simonsohn, Eric-Jan Wagenmakers, Jennifer J. Ware, and John P. A. Ioannidis. 2017. “A Manifesto for Reproducible Science.” *Nature Human Behaviour* 1 (1): 0021. <https://doi.org/10.1038/s41562-016-0021>.
- Nosek, Brian A, and Timothy M Errington. 2020. “The Best Time to Argue about What a Replication Means? Before You Do It.” Nature Publishing Group.
- O’Neil, Cathy. 2016. *Weapons of Math Destruction*. 1st ed. Crown.
- Parsons, Nick R, Charlotte L Price, Richard Hiskens, Juul Achten, and Matthew L Costa. 2012. “An Evaluation of the Quality of Statistical Design and Analysis of Published Medical Research: Results from a Systematic Survey of General Orthopaedic Journals.” *BMC Medical Research Methodology* 12 (1). <https://doi.org/10.1186/1471-2288-12-60>.

- Rigden, Daniel J, and Xosé M Fernández. 2022. “The 2022 Nucleic Acids Research Database Issue and the Online Molecular Biology Database Collection.” *Nucleic Acids Research* 50 (D1): D1–10. <https://doi.org/10.1093/nar/gkab1195>.
- Ritchie, Stuart. 2020. *Science Fictions: Exposing Fraud, Bias, Negligence and Hype in Science*. Random House.
- Rodgers, Peter, and Andy Collings. 2021. “Reproducibility in Cancer Biology: What Have We Learned?” *Elife* 10: e75830. <https://elifesciences.org/articles/75830>.
- Smith, G. 2014. *Standard Deviations: Flawed Assumptions, Tortured Data, and Other Ways to Lie with Statistics*. Duckworth Overlook.
- Stark, Philip B., and Andrea Saltelli. 2018. “Cargo-Cult Statistics and Scientific Crisis.” *Significance* 15 (4): 40–43. <https://doi.org/10.1111/j.1740-9713.2018.01174.x>.
- Welch, H. Gilbert, Jonathan S. Skinner, Florian R. Schroeck, Weiping Zhou, and William C. Black. 2018. “Regional Variation of Computed Tomographic Imaging in the United States and the Risk of Nephrectomy.” *JAMA Internal Medicine* 178 (2): 221. <https://doi.org/10.1001/jamainternmed.2017.7508>.
- Wernham, Ellie, Jason Gurney, James Stanley, Lis Ellison-Loschmann, and Diana Sarfati. 2016. “A Comparison of Midwife-Led and Medical-Led Models of Care and Their Relationship to Adverse Fetal and Neonatal Outcomes: A Retrospective Cohort Study in New Zealand.” *PLOS Medicine* 13 (9): e1002134. <https://doi.org/10.1371/journal.pmed.1002134>.