

Misleading Statistical Studies

Amanda Wolf

Misleading: That [which] leads someone astray, or causes someone to have an incorrect impression or belief; deceptive, delusive. (Oxford English Dictionary, online edition)

Introduction

Large statistical studies in the social sciences, including one-off or repeated cross-sectional surveys, time-series surveys and cohort longitudinal research, offer important numeric evidence for policy making. Although single studies rarely occasion dramatic policy shifts, statistical research findings can affect policy debate, even if not always directly or openly. At best, these studies reveal shapes and patterns in the social fabric relevant to health, safety, education and other social goals. Numerical measures of many social phenomena, such as unreported crime, illicit drug use, child-rearing practices or family composition, enter into a policy-making milieu crowded with competing numbers and qualitative information, as well as non-evidential values and power-based influences.

Statistical studies can cost millions, and draw significantly from modest research budgets. By one estimate, 39 public sector organisations planned to spend \$87.4 million on policy research in 2006/7 (MoRST, 2007). I have been told that a longitudinal study of New Zealand children and families, under development by a group led from the University of Auckland and the Ministry of Social Development, would likely have *annual* costs ranging from \$2 million to \$6 million or more. Getting the most public value from costly statistical research depends on how well researchers design their studies, and on how accurately and effectively they analyse and convey their findings to decision makers. In the process, the researchers themselves, academic and technical critics, the media, the public and government officials all interpret and apply numbers with more or less skill,

fortitude and scepticism. Accordingly, Joel Best, in his *Damned Lies and Statistics*, reminds us not to avoid statistics but to become ‘better judges of the numbers we encounter’ (Best, 2001, p.6). Better judgement means fewer *misinterpretations* of numbers, and ought, therefore, to result in fewer *misapplications* of numbers for policy purposes.

In this article, I leave aside misinterpretation and misuse of statistics, and their onus upon the ‘consumer’. Instead, I consider residual and insidious determinants of the public value of statistics: the ways in which good numbers, produced and reported by respected, unbiased, technically acclaimed researchers, may nevertheless *mislead* key actors in the policy debate. Misleading numbers can flourish even in the absence of sloppiness, self-interest or malign intent by a producer or advocate of numbers (matters which are well-addressed by Best). In the policy debate, a badly interpreted or misused number may be better than no number at all, because it can stimulate correction. But a misleading number may become embedded in the policy milieu with no further scrutiny. The less misleading a number is, the less will be the collateral damage from misinterpretation and misuse.

A comprehensive treatment of misleading statistics would require struggling through some roiling epistemological waters. Instead, the following briefly serves my purposes: no statistic perfectly reflects either the natural context of individuals and societies or changes in that reality. Rather, a statistic expresses the probability that a measurable quantity in a particular situation is really the case. ‘Reality’, or ‘truth’, has some qualities that may be known to an observer, but never beyond the shadow of doubt. At best, statistical observations are guesses. Therefore, even the best measures of reality may mislead, may cause the observer to have an incorrect belief about the truth. Statisticians have developed some

sophisticated techniques to minimise the proliferation of misleading numbers, and to accurately convey the uncertainty of numbers. Yet, outside their circle, among those who commission, fund and communicate research activities, this sophistication is often lost. Because no number exists in a vacuum, numbers can mislead if their initial presentation confuses people, or suggests wrong inferences to them.

In this article, I identify and illustrate several instances of misleading statistics, drawing upon four recent studies. I focus particularly on statistics used to ‘explain’ outcomes and to draw conclusions from comparative assessments. Such statistics attract keen scrutiny in government-sponsored research, especially when variables are tracked over time. Two of the four are products of New Zealand’s acclaimed longitudinal studies. One links abortion to mental health outcomes (Fergusson, Horwood & Ridder, 2006), and the other claims an association between work-related stress and depression and anxiety (University of Otago, 2007). The other studies are cross-sectional surveys. I selected reputable non-New Zealand studies that attracted entirely positive media attention: a study on the IQ advantages of first-borns, reported in *Science* (Kristensen & Bjerkedal, 2007), and a US Centers for Disease Control and Prevention survey that shows that men have more sex partners than women (Fryar et al., 2007).

All four are examples of large-scale, multi-million dollar, policy-relevant research. Some doubt that such studies actually *do* affect policy. But if they do not, one must query why not, given the resources invested. Frequently, of course, researchers seeking government funds must promise policy relevance. Perhaps understandably, they face incentives to assert more certainty in the evidence than is warranted. They are reluctant to note caveats. Within government departments, research is commissioned to investigate important social variables. Here, too, the incentives to collect and assess data with the highest policy relevance can, counter-intuitively, result in blind spots in the evidence base. I intend no critique of most of the measuring, categorising and analysing undertaken in the production and use of official statistics: these are nourishing waves replenishing our knowledge base.

The need to appreciate the causes of misleading statistics and to ameliorate their toll is especially acute for New Zealand. Large-scale statistical studies

are costly financially (even as they are relatively cost-effective in gathering and assessing masses of data). Yet the opportunity costs and flow-on effects may be many times more significant for the quality of available evidence and, ultimately, for social outcomes.

Four studies: what misleads and why?

Gina Kolata (2007), in ‘The myth, the math, the sex’, reports on a survey of sex practices, done under the auspices of the US Centers for Disease Control and Prevention (CDC), in which men claim they have had a median of seven partners in their lifetimes, whereas women claim four. The median calculations of ‘vaginal, oral or anal sex’ partners exclude people with no opposite sex partners (Fryer et al., 2007). Kolata quotes David Gale, an emeritus professor of mathematics at the University of California, Berkeley: ‘Surveys and studies to the contrary notwithstanding, the conclusion that men have substantially more sex partners than women is not and cannot be true for purely logical reasons.’

In principle, ‘sex partner’ is an objective, countable phenomenon. So, discrepancies in reported totals must be due either to gender-specific over- and/or under-reporting, or to gender-specific differences in the way ‘sex partner’ is construed. Unfortunately, we don’t know the degree to which either happens, nor the underlying mechanisms involved in construing and recalling events. While the survey design and analysis could be improved by better defining terms, or by further disaggregating the reported totals, the magnitude of the discrepancy is too large to be satisfactorily resolved by such methods.

Why was manifestly false data published with no reader advisory? The scientist’s answer, provided by the study’s lead researcher, is that data is what it is; the reasons why are not for her to gauge. Another answer, Kolata intimates, is that we accept what we expect: ‘[e]veryone knows men are promiscuous by nature. It’s part of the genetic strategy that evolved to help men spread their genes far and wide. The strategy is different for a woman, who has to go through so much just to have a baby and then nurture it. She is genetically programmed to want just one man who will stick with her and help raise their children.’ So, the numbers also mislead because they reinforce a stereotype, supposedly backed by rigorous science. Unfortunately, if policy makers need evidence on numbers of lifetime sex partners, this survey comes up empty.

The survey also collected data on illicit drug use. Should these drug-use statistics be used to plan enforcement and prevention activities? On the plus side, the data do not defy logic in the manner of the sex partner data. Against this, might drug users and non-users (the very same providers of the distorted sex partner data) contribute systematically inaccurate data? Similar concerns accompany *any* study in which the data are based on self-reports concerning activities and events that are conceptually vague: one person's recall and reporting of 'sex partner' or 'use of drugs' may not be another's.

The second illustration mainly avoids the limitations of self-reported data. Kristensen and Bjerkedal (2007) find that IQ falls as birth order rises (that is, children born second in a family will tend to have lower IQs than first-born children, and third-born children will tend to have even lower IQs). A younger sibling accrues an IQ advantage comparable to a first-born by becoming a 'social first-born' when an elder sibling dies in infancy. The difference amounts to a 'statistically significant' 2.3 IQ points between first- and second-borns. The data come from over 240,000 Norwegian 18- and 19-year-old male conscripts, who took an intelligence test as part of compulsory military service between 1985 and 2004. All conscripts within the specified period took the same test (though the year of conscription was controlled for in case the test was, by chance, easier or harder in one year than another).

Frank Sulloway (2007), in a companion comment in *Science*, is convinced 2.3 points matters, and finds a way to exemplify it. He writes, 'if Norway's educational system had only two colleges – a more prestigious institution for students with IQs above the mean, and a less desirable institution for all other students – an eldest child would be about 13% more likely that a secondborn to be admitted to the better institution'. Another reviewer, Roxanne Khamsi (2007), writes in *New Scientist*, 'The findings could suggest better ways of parenting the youngest children in a family' to overcome the 'social factors' that lead to their lower IQs.

Having assumed the finding's merits, Sulloway and Khamsi (whom I consider knowledgeable media exemplars) both fall into the trap of presenting (or implying) an explanation that fits, drawing authority from theories that do not feature in the study. Sulloway refers to a 'confluence model', which explains the

observation that older children tend to have lower IQs than younger children when tested as children, but then recover their first-born advantage by adulthood. The reasoning holds that older children score lower when they find themselves in a 'degraded' intellectual environment when a younger sibling arrives. Later, older children shift to the intelligence-enhancing role of tutor to the younger sibling. Sulloway admits there is little evidence for the tutoring effect. Khamsi, however, asserts the possible explanation that parents have more time and resources to invest in their first-born, but offers no backing for her claim.

Ever since Galton observed in 1874 that first-born sons were more likely than chance would predict to attain prominent positions, his negative association between intelligence and birth order has been confirmed in numerous tests using IQ measures. Yet just why this is so remained contested. Given no self-reporting bias or conceptual ambiguity, and with various statistical controls, the Norwegian study claims to support a 'family-interaction' theory rather than a competing 'gestation-order' theory.

Family interaction is a complex construct. As often happens in large statistical studies, the researchers adjusted the data, in this case for parental education, maternal age at birth, family size, birth weight and year of conscription. The report states, 'Because children from families with an adverse reproductive history had a less-advantageous distribution on a number of factors associated with low IQs, we considered it important to adjust for those factors.' In other words, the researchers statistically isolated 'family interaction' from some family-related events but not others (such as time spent with a child or sibling tutoring), to more accurately account for the association between birth order and IQ.

As variables, birth order and number of living older siblings have a conceptual clarity and precision that IQ scores lack. IQ *scores* are precise (given as 103 or 92, for instance), but the score's relation to 'intelligence' is not. The measurement of intelligence generally and the interaction of intelligence, IQ testing and social factors in the home did not feature in the small number of news reports and blogs I encountered. Yet Sulloway reports a consistent finding that first-borns are perceived as 'achievers' within the family. (Subsequent-borns occupy niches such as 'sporty' or 'clown'.) Might niche correlate

with test scores? Perhaps first-borns learn to be just a tad more serious about completing multiple-choice tests well. In short, there might be some *other* explanation that would explain the IQ findings.

My last two illustrations come from longitudinal cohort research. Such studies complement large-scale, one-off or repeated-snapshot studies. Rather than measuring a few variables in a very large sample, longitudinal studies measure a large number of variables a moderate number of times in a moderately sized sample. The data collection supports a wide range of discrete analyses. Hundreds of peer-reviewed papers have appeared from New Zealand's studies. Among these, for example, a reviewer of this article points out that a paper linking breastfeeding to academic achievement (Horwood & Fergusson, 1998) continues to be influential in scientific research. (Are first-borns perhaps breastfed longer than subsequent-borns?)

Recent analysis from the Dunedin Multidisciplinary Health and Development Study (University of Otago, 2007), which has followed 1,000 people since their birth in 1972/73, finds that 'work stress precipitates depression and anxiety in young working women and men'. High psychological demands, such as long hours, tight deadlines or pressure from supervisors, are associated with clinical depression, anxiety or both in women and men. At age 32, 50 women and 52 men were diagnosed with depression, anxiety or both for the first time. Because of the research design, researchers claim to have linked workplace stress to mental health disorders independently of other factors known to predict disorders, such as personality and socio-economic status. The misleading potential in this case arises from a combination of self-reported data, researcher involvement in setting thresholds for defining participants in or out of conceptually fuzzy categories, and the tendency to abstract selected associations from participants' lived experiences.

Longitudinal researchers invest considerable time gathering and analysing data. At each measurement stage, participants provide information and submit to tests over the course of a day. Corroborating evidence may be sought through a variety of means. Unlike numbers of sex partners and IQ scores, workplace stress levels and psychiatric disorders must be *inferred* by the researchers from the information reported by the subjects or measurements made of them. While

standardised diagnostic tools assist with identifying psychiatric disorders, gauging workplace stress is a matter of the researchers' qualitative interpretations of study members' self-reports. The authors defend their work: 'Other research has shown that self-reports of workplace stress are probably more accurate than reports by co-workers or supervisors. In the area of mental health, individuals' perceptions of their work environment are thought to be especially important.'

Yet they also write, 'workplace stress levels and psychiatric disorders were tested at the same time, [so] it is possible that depression may have influenced the answers given about work characteristics. The researchers did control for "negative reporting style" to account for this possibility.' That is, the researchers applied their *own* measure of reporting style to 'control' for a possible confounding variable. Since only 45% of newly diagnosed cases (46 individuals) were directly attributable to job demands, the effect of controlling for reporting style is ambiguous. What of the remaining 55%, for whom the 'association' between stress and depression cannot be directly attributed? What, indeed, of the 900 or so others, for whom any workplace stress or mental health concerns were judged below relevant thresholds, or whose work and mental health profiles otherwise differed? Nevertheless, the lead author asserts in a press release that '[i]n their 30s, most people are settling into careers, but it is also a time when people are at elevated risk for psychiatric disorders. Putting preventive efforts into reducing work stress at that age could bring big benefits' (University of Otago, 2007). Clearly, the author is directing readers to draw certain inferences from the numbers. But if no one questions the inference – are *most* people in their 30s *settling into careers?* – the numbers, whatever they show, are more likely to mislead.

The Christchurch Health and Development Study, New Zealand's other long-standing longitudinal study, has also been following a birth cohort over time. Recently, as the study participants reached 25 years of age, the researchers investigated various associations with abortion. Unlike previous studies, the Christchurch study compared women who have abortions with both women who had been pregnant but did not have an abortion *and* women who had not been pregnant. While adverse mental health effects from abortion have been attributed to guilt and unresolved loss, the

researchers expected that other ('third' or 'confounding') effects could be associated with both abortion and mental health outcomes. They find that 'those having an abortion had elevated rates of subsequent mental health problems including depression, anxiety, suicidal behaviours and substance use disorders. This association persisted after adjustment for confounding factors' (Fergusson et al., 2006). The research adjusts for 19 possible confounding socio-economic factors, childhood- and family-related factors, and health and personality factors. In addition, however, the researchers acknowledge a vague category of 'woman's circumstances at the time of the pregnancy'. The circumstances include her age, whether her pregnancy was planned, and the stability of her partnership.

Although the researchers appropriately note possible limitations in their study from omitted covariates and possible distortions due to respondents' under-reporting abortions, the catch-all 'circumstances' category points to an additional source of misleading information. The authors essentially admit that they do not know much about what influenced a woman in their study to seek an abortion (or not). Perhaps, they suggest, the mental health effects are due to the unwanted pregnancy and not to the abortion. No doubt many contextual factors, singly and in combination, are linked to mental health. In short, if context matters – where context is *a combination of factors* – neither modelling more variables nor improving the reporting of abortion incidence will add substantially to understanding.

Minimising misleading statistics

The illustrations together highlight three important ways that statistical studies can mislead. First, key constructs are open to wide interpretation regardless of whether data are self-reported, objectively measured and/or researcher-adjusted. Second, initial interpretations may overreach due to ambiguity in the variables used to measure complex social life. Ambiguity may arise from unquestioned consensus, or because researchers control for only some alternative associations, or because the notionally small step between a relatively unambiguous quantitative measure and some plausible conclusion actually traverses a deep chasm that invalidates the causal logic. Third, at the heart of the matter lurks a black box, the fuzzy and multi-faceted 'context' or 'personal factors' or

'subjective meaning', which cast doubt on both simple counts and causal explanations.

Plainly, studies of individual and social outcomes can *never* include all the contributing factors or possible explanations. The possibility that statistics will mislead cannot be avoided. Where does this leave us? Two recent books, one by William Starbuck, a noted management thinker with a background in science and engineering (2006), and the other a provocative argument by the Danish planner/geographer Bent Flyvbjerg (2001), struck me as particularly trenchant in their observations bearing on the mitigation of misleading statistics.

Starbuck claims that 'signals' and 'noise' look remarkably similar in statistical studies. Both occur as 'systemic components' and both vary erratically. While technical procedures attempt to distinguish the two, some mightily huge assumptions must hold for us to trust the results. Yet, instead of presenting hedged claims, researchers trumpet 'statistically significant but meaningless noise ... [and] often mistake confounding background relationships for theoretically important information' (Starbuck, 2006, pp.47-9).

Moreover, Starbuck claims, 'knowledge is what people say it is social processes elevate perceptions into facts, convert beliefs into truths' (2006, p.75). Statistics become the truths they estimate: first-borns *are* clever; men *do* have more sex partners than women. Flyvbjerg, however, drawing on Giddens' 'double hermeneutic', argues that what is to count as a relevant fact 'is determined by both the researchers' interpretations and by the interpretations of the people whom the researchers study'. He follows with the consequent implication: 'this means that the study of society can only be as stable as the self-interpretations of the individuals studied. And inasmuch as these interpretations are not constant, the study of society cannot be stable either' (2001, p.33): we *say* (for now) that first-borns are clever, we *believe* men have more partners (but 'sex' for men is not the same as it is for women).

Starbuck recommends that researchers *disturb themselves*, shake themselves out of their tendency to fall victim to poor assumptions. Then, he suggests, they should actively *experiment*, eschewing the classic strategies of building on previous research by more finely testing existing explanations or searching for something new in some overlooked crack in the landscape.

As Starbuck sees it, ‘There are many more combinations of symptoms than there are diagnoses, so translating symptoms into diagnoses discards information. Moreover, there are many more treatments than diagnoses, so basing treatments on diagnoses injects random errors. Doctors can make more dependable links between symptoms and treatments if they leave diagnoses out of the chain’ (2006, pp.108-9). He continues:

Academic research is trying to follow a model like that taught in medical schools. Scientists are translating data into theories, and promising to develop prescriptions from the theories. Data are like symptoms, theories are like diagnoses, and prescriptions like treatments ... Theories do not capture all the information in data, and they do not determine prescriptions uniquely The systems social scientists are trying to understand are very complex and flexible, perhaps too complex and flexible for traditional research methods that rely on spontaneous data and static analyses. (Starbuck, 2006, p.113)

Flyvbjerg adds context back into the picture, not because context holds variables that need to be brought under analytic control, but because context is *interpretively meaningful* experience:

The problem in the study of human activity is that every attempt at a context-free definition of an action, that is, a definition based on abstract rules or laws, will not necessarily accord with the pragmatic way an action is designed by the actors in a concrete social situation. Social scientists do not have a theory (rules and laws) for how the people they study determine what counts as an action ... because theory – by definition – presupposes context-independence. (Flyvbjerg, 2001, p.42)

His prescription, based on Dreyfus’ model of learning, reminds researchers that lower-level analytic cognition must be complemented by intuition and judgement at higher levels. Research must move beyond methodological formalism, just as Starbuck’s ‘doctors’ should leave diagnoses out of the chain. The goal is improved social dialogue on the questions of: Where are we going? Is this desirable? Who gains and who loses? The whole point, he says, ‘is to enter into a dialogue

with individuals and society and . . . to make moral debate part of public life’ (p.63).

Conclusion

We do not inhabit a 1984 world of immoral policy experiment. Nor do we choose to debate morality at every turn. Yet public policies, such as legalised prostitution or ACC coverage for workplace stress, *are* experiments. Values *are* close to their core. Policy-relevant statistical studies provide qualified measures of the facts of matters such as sexual practices and causes or outcomes of workplace stress. Statistics, and their initial presentation, carry with them the possibility to mislead. Starbuck and Flyvbjerg offer intelligent, complementary strategies that researchers can use (and that funders can support) to reduce misleading numbers. Researchers can adopt an open and exploratory stance by disturbing themselves and actively pursuing knowledge through experiment. And they can adopt an expert’s stance in judging contextual complexity. In the remainder of this article, I offer some suggestions along these lines. In the space allowed, I cannot expand each to a complete argument, so I must trust readers to think laterally from their own perspectives, aided by the preceding illustrations. Although the illustrations are big, expensive studies, in posing the suggestions I have in mind moderately sized studies, achievable by small teams.

New Zealand is a small and diverse country, where signals can be especially hard to detect in noise: Why *do* families form and dissolve? *Are* first-born children breastfed longer? It is wasteful to sift through masses of data and to accumulate associations in the hope that some will prove useful. Researchers should design studies to search for bigger needles in smaller haystacks. A big needle is a finding that is neither empty nor misleading. A small haystack is a search field selected according to specific criteria by a researcher who knows the contours of the *social* terrain, as well as those of the academic terrain. Time and effort are needed in the scoping stages of new research to develop and test variables and their measurement realistically and to craft focused data collection.

To confront vagueness and ambiguity in the scoping phase, researchers should disturb themselves by questioning measures of convenience or convention that might otherwise clutter a survey; by openly tracking their

assumptions as a way to find alternative measures of their main constructs; and by retaining multiple measures, not all of which are 'obvious', congenial to the researcher or even mutually compatible. Qualitative research, especially evaluations of prior policy 'experiments' and investigations of meaning in self-reported data, can direct researchers' attention to the more informative variables and associations, including some which will not yet have occurred to them. Researchers should confidently draw on their own knowledge and expertise, because these supply hunches that can help them to bridge weaknesses in formalised knowledge.

If richness of variables and measures intensifies, then scale may need to be reduced to maintain research tractability. Yet, a study with five freshly developed variations involving 100 participants each may be more informative than a study with 1,000 participants, which differs little from previous work. Variations can be approached serially, as the successive probes of researcher-as-'experimenter'. Whereas Starbuck's 'experiment' is tinged with social engineering – he favours introducing real changes to observe their effects – I adopt an interpretation that implies multiple probes of 'reality' in order to narrow uncertainty without oversimplification.

No number should be allowed to speak for itself. Too often we fail to adequately analyse, contextualise and interpret the data we collect. Conversely, some data are collected that can be only summarised, but not fruitfully analysed, in part because New Zealand's size and diversity limit the statistical power of tests. I would like to see researchers more actively engaged as expert spokespeople for – and against – various plausible interpretations of their findings. More surgically precise data collection, coupled with greater sophistication in analysis and comparison with findings from related research, could substitute for an undisciplined tendency to supply explanations beyond what the data can support.

Initial interpretations of research findings should be far-ranging – what's new, not new, missing, possible, surprising, disturbing, confusing? Findings should be examined in light of researchers' multiple activities to combat ambiguity and in light of prevailing 'truths'. Pursuing knowledge through experiment requires matching emerging findings to as many different situations and possible explanations as possible – what's the same, what's different? Applying judgement requires

the researcher to come out from behind an academic screen and speak to the public directly: 'I have immersed myself professionally in this field and this is what I see in this number.' Researchers should explicitly address prominent 'myths' when reporting results and openly discuss manifestly questionable data.

Interpretive ambiguity cannot be addressed at the level of semantics or statistics alone. Nor should consensus always be expected or demanded. Not every social phenomenon has one or a few generalised descriptions or causes that can be derived from a few variables. Thus, research designs should allow for multiple interpretations and draw on different individual and social meanings that are developed from a range of perspectives, rather than from expert theory alone. What level of commitment, or attraction, makes a person a 'sex partner'? What might account for the observation that one person avoids depression when a similarly situated person does not?

'Being critical', says Best, 'means more than simply pointing to the flaws in a statistic' (2001). In this article, I chose examples that contain excellent statistics. My aim was to direct attention to 'flaws' that mislead and that can be addressed by brave and open choices, especially at the design and initial reporting stages. Such added attention of this sort does not obviate the need to improve in other areas: we still need, for instance, solid baseline statistics and a more numerate media. We still need to counter an overreliance on simplistic analysis with both greater technical skill and creative qualitative analysis.

References

- Best, J. (2001) *Damned Lies and Statistics: untangling numbers from the media, politicians, and activists*, Berkeley: University of California Press
- Fergusson, D.M., L.J. Horwood & E.M. Ridder (2006) 'Abortion in young women and subsequent mental health', *Journal of Child Psychology and Psychiatry*, 47 (1), pp.16-24
- Flyvbjerg, B. (2001) *Making Social Science Matter: why social inquiry fails and how it can succeed again*, Cambridge: Cambridge University Press
- Fryar, C.D., R. Hirsch, K.S. Porter, B. Kottiri, D.J. Brody & T. Louis (2007) 'Drug use and sexual behavior reported by adults 1999–2002', *Advance Data from Vital*

and *Health Statistics*, 384, Hyattsville, MD: National Center for Health Statistics, 28 June

Horwood, J. & D.M. Fergusson (1998) 'Breastfeeding and later cognitive and academic outcomes', *Pediatrics*, 101 (January), <http://pediatrics.aappublications.org/cgi/content/full/101/1/e9>

Khamsi, R. (2007), 'Firstborn children are the cleverest', *New Scientist*, 21 June

Kolata, G. (2007) 'The myth, the math, the sex', *New York Times*, 12 August

Kristensen, P. & T. Bjerkedal (2007) 'Explaining the relation between birth order and intelligence', *Science*, 316, p.1717

Ministry of Research Science and Technology (MoRST) (2007) *Public Sector Financing of Research 2006/07*, Wellington: MoRST, <http://www.morst.govt.nz/Documents/publications/statistics/Public-Sector-Financing-of-Research-2006-07.pdf>

Starbuck, W.H. (2006) *The Production of Knowledge: the challenge of social science research*, Oxford: Oxford University Press

Sulloway, FJ. (2007) 'Birth order and intelligence', *Science*, 316, pp.1711-2

University of Otago (2007) 'High pressure jobs linked to depression and anxiety', press release', http://www.otago.ac.nz/news/news/2007/01-08-07_press_release.html

Amanda Wolf is a senior lecturer in the School of Government at Victoria University of Wellington. Her main research interests include methodologies for policy research, information-based policies and subjectivity in policy contexts.

FREE AND FRANK: Making the New Zealand Official Information Act 1982 work better by Nicola White

The latest Institute of Policy Studies publication

The New Zealand Official Information Act 1982 is frequently hailed as one of this country's most significant constitutional reforms. It is praised as world-leading in its refusal to contemplate that any category of government information might be completely immune from the prospect of public disclosure.

But for those who work with the Act, either as seekers of information or as officials responding to requests, it is not an unqualified good - the working reality of the Act can be frustrating and time-consuming, just as often as it is enlightening.

This book follows a two-year research project into the day-to-day operation of the Act. It examines the history of the Act's passage and subsequent development and reports the candid views of (anonymised) officials, politicians, academics, political advisers and 'regular seekers' of Official Information.

The result is a 'free and frank' picture of the operation of administrative and political processes around Official Information. It carries a sobering message about how those processes, if not well-managed, can erode trust in government across time.

But the picture is not all bleak. The book concludes with proposals for change designed to build on the very real strengths of the system, to help it cope with the challenges of modern politics and the information age.

Nicola White has had nearly 20 years of close involvement with the theory and practice of the Official Information Act as a public sector lawyer and policy adviser. This research was carried out whilst she was a Senior Research Fellow at the Institute of Policy Studies in Victoria University's School of Government (2004-6).

Published – November 2007

Format – B5 Paperback, pp 318

ISBN – 1-877347-21-3

Price – \$35.00 (including P&P within New Zealand)

To have a copy of *Free and Frank: Making the New Zealand Official Information Act 1982 work better* and an invoice sent to you, please email, phone, fax or mail your order to:

Institute of Policy Studies

Victoria University of Wellington

Email ipos@vuw.ac.nz

Telephone +64 4 463 5307

Fax +64 4 463 7413

PO Box 600, Wellington

New Zealand