

# An evidence spectrum for social services policy and practice that builds trust<sup>1</sup>

Len Cook<sup>2</sup>

Revised<sup>3</sup> 18 January 2017

## Abstract

Social services face great demands not only in improving the condition of society, but even when ameliorating or making manageable chronic complex situations. Accountability mechanisms abound, incentivising risk aversion while seeking innovation and efficiency. Social policy choices are inevitably better informed by relevant evidence, but time pressures, fiscal and political compromise and insufficient capability often bedevil the capacity for effective deliberation. Targeting has been a critical instrument of social policy for some three decades, with social investment adding a new dimension to targeting that suggests that social policy may be at a crossroads. The evidence that politicians and institutions choose to gather and the models that they use will predetermine policy choices. The evidence spectrum provides a basis for understanding where this is happening.

Once policy has been determined, putting it in place will be more fraught the weaker the evidence base behind it which can inform implementation. Trust in policy is tested the most at the point of delivery, and the costs of poor policy are borne by the recipients. No matter how well developed, the evidence at the time of policy development will be an incomplete source of the knowledge needed for effective implementation. Once implemented, knowledge from process evaluation and continuous improvement can accumulate and modify processes, but only if they are able to occur. Institutional cultures and political mind-sets can prevent findings being produced which challenge the quality of programmes. Independent validation of the quality of services is vital to limit this occurring.

The rich expansion of information becoming available is challenging long held beliefs about privacy, information ownership and ethical practices. Citizens pay a high cost when the opportunities are not taken from this expansion. At a national level, information policy needs to reflect these new challenges, which are only partly acknowledged by the mantra of 'evidence based policy'. This paper brings together these diverse issues through its concept of an evidence spectrum, where sources, policies and structures all could have more coherence than seen currently in public administration in New Zealand.

---

<sup>1</sup> Initially presented at a Symposium on the retirement of Associate Professor John Harraway, University of Otago, September 2017

<sup>2</sup> Len Cook was the Government Statistician of New Zealand from 1992-2000 and National Statistician of the United Kingdom 2000-2005.

<sup>3</sup> I am grateful to Vince Galvin and Dr Bob Stephens for comments on early drafts which changed the scope and form of the paper

## Table of contents

<b>Introduction</b> .....	3
<b>Politics and policy – direct and indirect influence on practice</b> .....	4
Practicalities of government decision-making .....	4
The political nature of social services .....	5
Political and practical dimensions of social investment .....	6
<b>Public trust</b> .....	8
The limits to consumer power in social services .....	9
Testing fitness for purpose.....	10
The impact of targeting on trust.....	10
<b>Evidence and policy</b> .....	11
The selection of evidence and policy.....	11
The broad spectrum of evidence potentially informing social services.....	12
Scientific capability .....	14
The significance of integrability on the scope of the evidence base .....	15
<b>Information forms and the evidence base</b> .....	16
Official statistics and the questions we need to ask.....	16
A place for evaluation .....	16
Integrated Data Infrastructure .....	17
Progressing towards a de facto population register.....	18
Modelling .....	19
<b>Leadership of social services structure and design</b> .....	20
Opportunities from system-wide leadership in the evidence base.....	20
Evidence that services work - continuous improvement.....	22
<b>Māori</b> .....	23
<b>The place of uncertainty in informing social services policy, delivery and structure</b> .....	24
The many natural sources of uncertainty .....	24
Pacific .....	25
An indicative measure of selection bias.....	25
<b>Conclusion</b> .....	26
Key themes.....	27
Wisdom on trustworthiness - The philosopher, statistician, poet and politician .....	28
<b>APPENDIX: Opportunities and constraints in the evidence base for social policy and services</b> .....	29

## Introduction

---

For some two decades now, ideas such as ‘evidence based policy’ and ‘evidence informed policy’ have signalled a political commitment to strengthening evidence gathering and evaluation capability across government. In the past, the variability and risk aversion of political and institutional decision-making has provided a constraint on our willingness to learn from available knowledge, or create new knowledge, compared to most other areas of public policy. Non-use or misuse of evidence brings huge costs to citizens, by crowding out valuable alternatives, as well as increasing the cost that citizens incur in order to get the best they can from services that could have been better designed. The term ‘evidence based policy’ itself risks narrowing recognition of the inevitable interplay between politics and policy. Politics not only shapes policy, but influences the different roles in government of politicians and of the public service, limits the broad scope of evidence and its uses, and enhances the influence of more subjective information forms. Politics and institutional cultures shape the institutional structures, obscure the importance of independent validation for ensuring trustworthiness, as well as oversimplifying how public trust in information issues is enabled.

Whilst policy choices will inevitably be political, operational decisions ought to be protected from political influence by the structures and impartiality<sup>4</sup> of the public servant, and overseen by independent oversight. In this paper, the spectrum of evidence that is potentially available to bring about the trustworthiness of social services policy and practice is analysed.

There has more recently been a major expansion of observational data from the integration of records from various administrative processes, which so far has proceeded ahead of commensurate increases in analytical capability for improving the social services. This has triggered a thirst for better use of evidence around the sector, among NGOs as well as government. While the sheer diversity, complexity and variability of people puts limits on the precision of measures and models, the capacity to draw on people’s experiences to put a window on parts of the life course of people has magnified, enabling some of the opportunity costs of any past policy decisions to be made transparent, including those based on anecdote, ideology, untested theory and attitudes.

Fundamental limitations to the quality of social services will persist, but through a wide mix of capability initiatives the resources to manage some of them are expanding. Steps that take account of the qualitative and quantitative knowledge we can have about the opportunities and limitations of measuring, summarising and modelling many aspects of the human condition will help inform these limitations. In a time of rapid change in the evidence base, those involved in social services policy and service delivery need to be able to understand how to take advantage of the potential of the data sources, and effect their integration. Models influence policy choices, and where models have weak statistical validity, the mechanisms to have this made transparent are poor. We need to make clear how citizens can assess the trustworthiness of the methods and analyses behind any policy, and ensure that those affected have reason to trust the sources and analytical methods used on these enriched evidence forms. Without a robust quality management framework, the uncertainty inherent in all information can be easily ignored in decision-making about policy, in service delivery or in establishing the structure of the social services system itself. Without such a framework, we will not know the comparative strength of the many forms of observational evidence we might have, especially when competing against alternative sources of information from anecdote, ideology or belief. A possible obligatory quality management framework for evidence and its use that impacts on people’s lives is drawn from attributes

---

<sup>4</sup> Bob Gregory. “A Ride on the Ridgeway Bus”. Policy Quarterly Vol 13 August 2017

of randomisation, representativeness and realism that underpin statistical practice. Given that evidence forms that have high statistical integrity are often not available or can fit the issue in hand, without knowing the comparative strength of other forms of evidence it can often be ignored because its validity is not clear.

## **Politics and policy – direct and indirect influence on practice**

---

### **Practicalities of government decision-making**

Governments exist to govern, not only by putting in place their election promises, but also responding to emerging situations. Governments must make decisions, and how they prepare for this has an impact on the scope and quality of services that citizens can receive. Governments can draw on a wide range of analytical expertise along with information sources including official statistics and existing research resources. For many issues, it may not be enough to understand the context, evaluate target populations and compare policy options. The perceived or actual imperatives of timely decision-making mean that governments will often not be able to initiate evidence relevant to the issue at hand. Social services that have been developed without an appropriate evidence base cannot assess and keep in check the effectiveness and efficacy of service delivery. Experience in evaluating social services programmes would accumulate knowledge about the likely effectiveness and efficacy of alternative service delivery forms, which may facilitate cost benefit analyses and policy comparisons. In situations where the quality of the evidence base is recognised as not strong, the evidence obtained operationally by delivery staff and their autonomy on the ground must be expected to play a larger part in the effectiveness and efficacy of the policy. Where the quality of evidence in both policy formation and operational practice is poor, then citizens bear the risks of policy failure through both excessive fiscal costs and the personal costs and risks they bear as service recipients. Regardless of the foundations of any policy, its implementation needs regular monitoring to enable continual adaptation to change in the context of the service, and continuous improvement. As the scope and connectedness of the evidence base increases, the opportunity cost to citizens from policy that is made in ignorance of such an evidence base or avoidance of it will become more apparent. Tests of randomness, representativeness and realism of both information sources and models underpin assessments of quality.

If we wish to understand why, and for whom else what we do might work, or whether what works now will continue to do so, then scientific methods of gathering and analysing evidence are essential. Initiating this is made harder where there are very diverse motives in the commitment of politicians and institutions about accumulating and using evidence. This makes it difficult for citizens to then work out if they can trust public policy decision making. When rare<sup>5</sup> adverse events arise, their occurrence can trigger fundamental change in policy or practice. Much established practice that draws on experience is not well codified. This makes it difficult to transfer knowledge outside of apprenticeship type processes for learning on the job, and it also understates the contribution to decision making in practice of accumulated experience. This experience will be discounted where the autonomy of frontline staff is constrained.

When we have implicit or explicit constraints from political or institutional processes on what can be evaluated, the higher the opportunity costs borne by citizens from the lack of a comprehensive testing and monitoring framework of the services that result. When the costs of poor service are excessive

---

<sup>5</sup> CRI-2015-085-002309 [2016] NZDC 12806 Reserved judgment of Chief Judge Jan-Marie Doogue

compared to the benefits from services that are effective, evaluation of service design needs to be founded on scientific knowledge that can be subjected to strong scrutiny. Knowledge of the reliability is a vital part of the evidence. There are methods for the qualitative and quantitative assessment of the quality of evidence. Ignoring the reliability of evidence is to willingly remain ignorant of its fitness for purpose and the practical limits to quality that result from political, cultural and institutional restraints, or the need for personal safety. Operation complexity and operational failures limit what we know, and many variables are prone to respondent or agent variability. Diversity in the human condition can bring about selection bias, lead to incomplete/partial description of characteristics, and generate variation that remains untested or unaccounted for. Model misspecification that can lead to the spurious estimation of attributes will exacerbate any existing selection bias, and the impact will be dependent on the nature of any negative consequences on those who are falsely selected (false positives), and on those who should be selected but are not (false negatives).

### **The political nature of social services**

Social services include a wide-ranging mix of income transfers, care and protection services, developmental activities and remedial actions. Despite the resources used, and the potential for both harm and good that can be done, we often do not have available an analytical basis for determining why most programmes take the form that they do, and how well they work. Some of this is because of the nature of social services while much is because the past choice of social services mainly reflects elements of social and cultural values and preferences that make them immensely political. Many social services involve partly connected networks of groups and agencies, often operating with some form of contract to a government agency, with partial information about the people that they are helping, and with separate and potentially different ways of defining people and their conditions, and establishing needs.

There are fundamental elements of social services that result from politically determined rather than scientifically based perspectives on the nature of programmes: – universal versus targeted, service provision versus cash, free or part charges, outsourced or public provision, and the nature of the emphasis on human rights. Philosophical attitudes to taxation, regulation, penalties and sanctions also shape political preferences. Certainty in the means can occur without clarity in the results sought. Whatever the political perspective, evidence is essential to improve the quality and robustness of service outcomes over the time periods for which costs and benefits are to be compared. There needs to be the ability to take account of the relative impact of dispersed contributors to improved outcomes, and of the breadth of outcomes considered. Even so, the selection of social services is a primarily a political action, and the imperatives of deciding what to do will rarely await the discovery or implementation of new information sources. At its crudest, programmes will be judged by how much funding will be made available for them rather than what they will achieve, as typically occurs at the time of general elections. In the most recent general election in New Zealand, policy commitments were generally expressed as ‘how much’ rather than ‘how useful’. One consequence of serious concern is that once decisions have been taken, the political risk that previously unavailable information might adversely challenge the original decision can result in the prevention of improved information sources that might better inform future situations, and the quality of services that are provided.

The expectations on social policy and service delivery are continually being extended. The rise in importance of a social concern generally occurs without any formal process, and often it is the community sector that triggers interest in concerns that get addressed at a government level. Issues

that are currently in various emergent stages of wider recognition in the policy information system include: obesity, social media and technology, super diversity, suicide, harm and violence, abuse, bullying, mental health, drugs, pornography, antibiotic resistance, homelessness, crowding, fertility, urban infrastructure, incarceration costs and third world diseases. Where we can develop an evidence base to inform a political or institutional response, the long-term viability and impact of that response on citizens will reflect the limits to measurement and analysis. Without knowledge of the evidence base there is likely to be unjustified confidence in services, and this can bring damaging consequences for groups of citizens.

In the social services, institutions and politicians appear more likely than in other sectors of public administration to have a strong aversion to evaluation and continuous improvement practices that make transparent the imperfection inherent in their decisions and complicate managing political risk. There are many examples where the worlds of science, policy and practice remain not as readily connected in the social services as might be presumed. It is not unusual for long-standing social services to have been at best poorly tested and evaluated, with the consequence that the final form of many programmes is not based on relevant evidence or regularly tested by replication, pilot studies or continuous improvement practices.

Observational evidence always competes with anecdote, belief<sup>6</sup>, un-validated theory or just prejudice, not just in setting policy but also in determining whether to continue to gather evidence or invest in new forms. Where evidence is underused through choice, limited competence or underinvestment, then anecdote, belief, un-validated theory or prejudice are more likely to drive decisions. The very existence of evidence that people have reason to trust can inform or challenge the political or institutional preferences that frame thinking and problem solving. There is an inevitable and potentially quantifiable uncertainty and system risk inherent in social services which are poorly understood by the public and by politicians, for example in decisions such as granting parole. Where we have little idea of the quality of the evidence used to justify the policy, we cannot take account of that uncertainty in managing the delivery of the service, such as the undoubted limits to the of precision of any screening criteria. Consequently, with many social services we are often unable to be certain about what happens for the people who have a need and entitlement for them, and why only some of these people connect effectively with a service that they are eligible to obtain.

## **Political and practical dimensions of social investment**

The contemporary focus on social investment has been to challenge much of the received wisdom about the quality of connection between departmental performance and the experiences of citizens, as measurable by effectiveness and efficacy. However social investment evolves, it will necessitate a major commitment to evaluation in its various forms. While social investment has been connected to an emphasis on the place of the market compared to government, this does not explicitly define or determine social investment. There are universal elements of:

- a focus on the long-term outcomes when making choices
- measuring performance needs to focus more on citizens than agencies
- investment in research and evaluation into interventions that contribute to knowledge of ‘what works’

---

<sup>6</sup> ‘Three strikes and you’re out’ example, and boot camp proposals

- a need to remedy weaknesses in the gathering, accumulation and use of evidence, and exploit the unrealised potential of data resources
- improve the ability to have an effective contest for resourcing need, care and support that has the most long-term benefit
- advance the ability to manage ethical and human rights issues as forms of data gathering and use evolve.

Political choices can take social investment in diverse directions, affecting the balance of universality and targeting, the balance between the market and government production of services, and belief about the capacity to individualise life course models when extended into the future. The strongly individually oriented focus for social investment developed in New Zealand by late 2017 had led to identifying additional elements:

- tightened eligibility tests with priority focus on those individuals deemed to be the most vulnerable, or the outcome of multigenerational conditions and adversity
- individuals to be targets of initiatives initiated by delivery agencies as selected by a risk assessment of individuals formed from analyses of group characteristics
- assessment of the level the future fiscal liability as determined by quantitative criteria selected to summarise the individuals current condition and potential
- the use of sanctions where citizens did not conform with eligibility or entitlement obligations
- a strong emphasis on assessing the fiscal compared to the social nature of long term benefits, involving the rethinking of evaluation of benefits and costs over a long time-period or lifetime
- a strong preference for models of delivery involving NGOs, commercial organisations, with public sector agencies only where necessary
- using the assessments of future fiscal liability, responsibility for specific individuals could be transferred to third parties through using social bonds as a way of creating financial incentives to maximise the potential for improvement of the individual
- accumulation by government of identified client transactions as a condition of funding for any purpose of that social sector NGO
- a strong analytics centre for applying a strong analytical capability with individual service providers (Social Investment Agency).

One consequence of the strongly individually focused approach to social investment is that the use of algorithms for screening has advanced more rapidly than any consequent independent validation and oversight. These algorithms appear more prevalent in parts of the social services and justice sector where there is no standing judicial oversight body, particularly in child protection, the Department of Corrections (Corrections) and the Ministry of Social Development (MSD). Expert independent oversight of methods is weak compared to the impact of selection errors on people's outcomes. This form of social investment tends to stigmatise targeted populations, reinforcing the two-tier system that now exists between universal services in health and New Zealand Superannuation, and the rest of the welfare state.

One alternative is a more universalist approach to social investment. A stronger political focus on universality and strengthening of public service capability would potentially shift to a different set of imperatives, that would most likely include a deeper understanding of the importance of great variety of

evidence than has been the case with social investment so far. This would more clearly determine the focus of those with responsibility for the management of issues of a cross agency character.

This needs to include:

- a foresight capability monitoring emerging issues to formulate early strategies to mitigate or reverse disturbing trends
- understanding of what families and whānau do by themselves in care, health, housing and education
- obliging agencies to carry out and public evaluation studies of their services:
  - strong focus on assessing cultural relevance in evaluation
  - more visibility and recognition of selection errors to focus process improvements
  - report regularly on the take-up by New Zealanders of the services that people are eligible for
  - strengthened continuous improvement practices
  - rethink focus of performance measures.
- maintain a critical mass of experts in evaluation who can work on long-term, cross-agency issues such as youth mental health and family violence
- maintain relationships with academic experts in fields relevant to social services through a body akin to the Superu Social Science Experts Panel.

An evidence based design of the social services system would place greater emphasis on the impact on outcomes for people served by the system. This would involve in part:

- using information from the Integrated Data Infrastructure (IDI) to identify service performance improvements
- a comprehensive rethinking of methods used to ration services
- a rethinking of the impacts and place of penalties and sanctions
- finding ways to identify cohort or community group effects in individual data, e.g. for suicide, self-harm
- effective evaluation of past system failures, e.g. institutional child abuse
- using information from the IDI about agency connections to redesign the scope of agency activities
- drawing on the knowledge obtained by NGOs in operating services.

## **Public trust**

---

The political dimension of policy is reflected in the statutes and political choices made by Ministers. The strongest check on this is through the electoral system, while along with Parliament itself, the judiciary provides assurance that the executive government of the day and the agents of the state can be held to account for their actions. Parliamentary officers, particularly the Comptroller and Auditor-General oversee the integrity of actions, particularly financial appropriations. Parliament is the supreme authority, but is constrained by the dominance of in the House of Representatives by Ministers or Ministers in waiting in New Zealand, who display little preference for a career as legislator unlike larger Houses of Parliament. The rule of law is given effect to in service delivery by a professional public service that has obligations to treat all citizens impartially in the statutes that they administer. The judiciary, and a variety of independent quasi-judicial bodies are often the strongest validation of the impartiality of any process or experience that is external to the organisation being challenged. The capacity for



judicial challenge has become more significant through the growing political intrusion into operational matters because of the 'no surprises' policy of successive governments, and the developments in public sector management of the late 1980s with the mantra 'the Minister is the client'. Sir William Harcourt, a nineteenth century Liberal English politician, summarised rather quaintly the roles of Ministers and civil servants when he said that: "The Minister exists to tell the Civil Servant what the Public will not stand".

The dominant focus of the performance management regime is on agency performance rather than the consequences for citizens, and this has enhanced the potential for impartiality to be under threat. Bob Gregory<sup>7</sup> expands on this development and draws on the work of Rothstein and Varraich in developing a connection between a loss of impartiality and corruption.

Baroness Onora O'Neill has argued<sup>8</sup> that a proliferation of accountability mechanisms by governments did not necessarily increase trust. She asks whether systems of accountability are meant to replace trust or to improve the basis for placing and refusing trust. She concluded that;

*"To be accountable is not merely to carry a range of tasks or obligations, for example to provide medical treatment to those in need, to make benefit payments to those entitled to them, or to keep proper accounts. It is also to carry a further range of second-order tasks and obligations to provide an account of or evidence of the standard to which those primary tasks and obligations are discharged, typically to third parties, and often to prescribed third parties."*

### **The limits to consumer power in social services**

That the consumers of social services have minimal consumer power (come-at-ability) means that their withdrawal of engagement would generally be unlikely to have any influence on improved treatment of future consumers. Complaints are rarely systematically recorded and reviewed, and independent oversight is not universal. Consequently, unless agencies have a continuous improvement programme or use operations research modelling, the often-belated connection of service components has been left to citizens, who must incur high transaction costs. These are rarely if ever considered when developing policy. The performance measures of government are generally focused on accountability for agency efficiency as opposed to how well they meet the demands of the client for obtaining desired outcomes. Where there are insufficient means to recognise whether the costs and risks of programmes are transferred from agents of the state to citizens, this will further limit the quality of the long-term outcomes of citizens in unaccountable but significant ways. Sanctions to incentivise compliance can be applied quite loosely<sup>9</sup> if there is no monitoring of their impact on consumers.

Where there are not strong processes for holding government and its agents to account, transparency about the quality of evidence is of much greater value to the targets of decision-making than to decision-makers in policy or service delivery. There are parallels outside of government for having a broad perspective on trustworthiness. These include the non-publication of adverse clinical trials in testing drugs, limiting the type of studies included in meta studies, and the silent funding by industry lobby groups of purportedly independent studies by professionals selected and funded by them.

---

<sup>7</sup> Bob Gregory. 'A Ride on the Ridgeway Bus'. Policy Quarterly Vol 13 August 2017

<sup>8</sup>Onora O'Neill, *Holding Accountability to Account*, Royal Statistical Society Beveridge Lecture 2009

<sup>9</sup> Housing New Zealand testing for methamphetamine wrongly led to people being forced out of rental accommodation. Poor testing and management of the town water supply in Havelock North led to people there experiencing the largest outbreak ever of campylobacter

Where the sources and dynamics of diversity and variability of the population are significant but not well understood, there is a risk that policy will be based on an unrealistic capacity to address its impact. Therefore, service delivery agents then have no option but to manage the unanticipated complexity in the population targeted by the policy. Without the operational autonomy to adapt eligibility and entitlements where justified, the introduction of the services may undermine rather than increase public confidence in the service delivery agents, and perhaps the social services generally.

### **Testing fitness for purpose**

In social policy and service delivery, testing for fitness for purpose can draw on traditional statistical inference. This paper suggests that while the political context for investing in evidence and selecting programmes makes it unlikely that a strong synergy between science, policy and practice will occur without active leadership, there are many other practical reasons why the synergy can be poor and services of doubtful quality. Statistical thinking has a critical part in assessing trustworthiness of any evidence base applied to policy decisions.

Methods of measurement, modelling and analysis affect the lives of groups of people, either through the application of policy at a macro level, or by determining the delivery of services at a personal level. Consequently, their quality needs to be made transparent by both qualitative and quantitative assessments. To ensure the integrity of such work, government urgently needs to develop an obligatory quality management framework for evidence and modelling that impacts on people's lives. In other fields where evidence has a major impact on the well-being of people, quality frameworks are well established, and are integral to the trustworthiness of the services. To be relevant for New Zealand, such a quality framework must recognise what is needed to be relevant to Māori, and make transparent how the distinct characteristics of other cultures are taken account of. Getting it wrong can adversely affect citizens as well as benefiting them when getting it right. Transparency through validation of an obligatory standard could be comparable to practices well used in official statistics. The concepts of the statistician (randomisation, representative, realism<sup>10</sup>) provide a strong and well tested foundation for determining how to make transparent just how far it is possible show how far to place trust in information that we do have, and enable trustworthiness of policy and services to be rightly anticipated by others.

### **The impact of targeting on trust**

In terms of the volume and variety of transactions, the social services sector is probably the most complex area of public administration. The level and nature of these transactions changed as targeting became more pervasive throughout the 1980s and 1990s. In retrospect, the shifting from universal programmes to increasingly intensive targeting during the last half of the 1980s could not have taken place at a more challenging time. Greater population diversity, multiple languages, an increased range of service options, increased prevalence<sup>11</sup> of disorders that require intensive support, greater longevity and larger numbers of the infirm aged, the consequences of synthetic drug addiction, heightened family formation and dissolution, greater recognition and responsibility for child abuse and family violence as

---

<sup>10</sup> Kish, L. (1987). *Statistical design for research*. New York: Wiley

<sup>11</sup> E.g. The number of children diagnosed with autism or related disorders has grown at what many call an alarming rate. In the 1970s and 1980s, about one out of every 2,000 children had autism. Today, the CDC estimates that one in 150 8-year-olds in the U.S. has an autism spectrum disorder, or ASD.

well as community treatment rather than institutionalisation of those with complex mental health conditions have all since required a degree of sensitivity to individual circumstance compared to the expectations of the universalist system that served the generations up to the post war baby boomers.

Some social services of the state have become increasingly targeted, and more focused on those with demonstrable complexity and vulnerability. Other state services, like many consumer services, have become de-personalised using call centres, internet access, less immediate and gated access to front-line staff, intimidating bureaucratic requirements and a narrowed view of core business. Family and community organisations act as intermediaries between the citizen and the state, and this is especially important for those deemed to be at risk clients, including ex-prisoners, people with mental illnesses and those from communities still subject to racial prejudice. Statutory services, through their codification of responses, expect and oblige applicants to meet predetermined characteristics and are unlikely to give front-line staff the autonomy to detect and respond to conditions that are not anticipated in service design. Not all information of importance is obtainable by the codification of information gathered through rules based processes. There are areas of high need for support, including domestic violence and sexual abuse<sup>12</sup>, where a large share of victims believe that they cannot trust others with their experiences. The integrated services pilots for family violence have demonstrated very large gains in acceptance, engagement and commitment for services that have focused first on the outcomes consumer. Were it not for the cost of such programmes, the models adopted have huge potential in ameliorating otherwise intransigent situations, and bringing alternative pathways to social services delivery.

Screening methods, including those informed by ‘predictive modelling’ approaches, are tools that are becoming prevalent in targeted services, yet we have done little to accumulate knowledge about their bias and variability, and consequent influence on the trust of those who experience them. Where there are services that involve several organisations in the delivery to the consumer, selection bias and variability can result in different priorities and responses. Similar criticisms apply to psychometric testing where the validity of the underlying models for different cultures, gender and generations is not transparent.

## Evidence and policy

---

### The selection of evidence and policy

There are ongoing tensions faced in all social policy decisions which could be influenced by knowledge of the quality of evidence and processes. The development time for new research and scientific deliberation can rarely match the urgency with which serious conditions emerge and need responding to. Having formal structures for evidence gathering and analysis challenges institutional inertia around historical commitments and investments, and risk management. New Zealanders value privacy, yet the new frontiers of evidence involve building up rich life course histories of individuals. The connection between transparency and trust is complex and uncertain. Trustworthiness is for the judgment of citizens, not agents.

---

<sup>12</sup> The Australian Royal Commission into Institutional Responses to Child Sexual Abuse found that the mean lag between experiencing abuse and telling someone the first time was 23 years.

The performance measures of public agencies are dominated by indicators which demonstrate the fiscal and efficiency achievements for that agency that are expected by the agency Minister. Some indicators are shared by several agencies. The measures generally are weak in signalling changes to the quality of services received by citizens as consumers, or the cost to consumers of engagement. Extending agency performance measures can reduce the service received by consumers and reduce the benefits from innovation and service improvements. Continuous improvement practices and evaluation studies are infrequent in the social services sector, and this has been highlighted by several reviews, the Productivity Commission<sup>13</sup>, Professor Sir Peter Gluckman<sup>14</sup>, and the Review of Expenditure on Policy Advice<sup>15</sup>. There is rarely just one form of evidence that would be sufficient to fully justify trust in any policy. Natural experiments that result for unplanned and undesirable situations such as earthquakes, water system failures, institutional failures such as child abuse or violence provide rich sources of information, should independent investigation take place.

When we try and solve complex problems, we can end up not using some knowledge that should have informed the design of solutions. It is only this decade that administrative records have been having the high value now recognised, but we have a legacy of poor information management across the public sector, with weak system-wide meta data standards to drive integration, and uncertainty about quality. The existing evidence base is underused through weak analytical capability<sup>16</sup>, and poor appreciation of the strengths and weaknesses. The way our politics and institutions operate means that performance measures generate information that is often unrelated to fundamental concerns. They have led to government interest in the experiences of community organisations being narrowed to fiscal accountability, while understanding of what families and whānau do is similarly weak. There is little understanding or taking account of cultural capital and cultural institutions until some explicit obligation is in place. Institutional and political culture can limit the scope of evidence gathering or encourage it. The next section of this paper discusses these issues in more depth.

How concerns are framed will determine what is needed by way of policy and evidence. Without deep insight, the available evidence can narrow how to frame concerns. Family violence had long been regarded as a private household matter until some fifteen years ago when the police began to treat perpetrators as criminal offenders justifying court action. More recently, family violence thinking has extended to recognising continuing harm, involving children as well as partners, while recognizing trauma as well as physical violence as outcomes needing treatment. Rehabilitating perpetrators is also recognised as a necessary response. The Havelock North water poisoning resulted from a failure of the Hasting City Council to recognise that its water management was an integral part of the public health service, and needed to be managed as such.

## **The broad spectrum of evidence potentially informing social services**

Without an evidence base built on scientific accumulation of observations of people and processes, and a commensurate analytical capability, policy and practice are more likely to reflect a subjective world view that can involve cultural, religious or social bias. Whenever political sentiment or institutional culture has been the dominant determinant in policy selection and programme delivery, the foundation

---

<sup>13</sup> NZ Productivity Commission 2015 'More Effective Social Services'

<sup>14</sup> Sir Peter Gluckman, September 2013: The role of evidence in policy formation and implementation A report from the Prime Minister's Chief Science Advisor

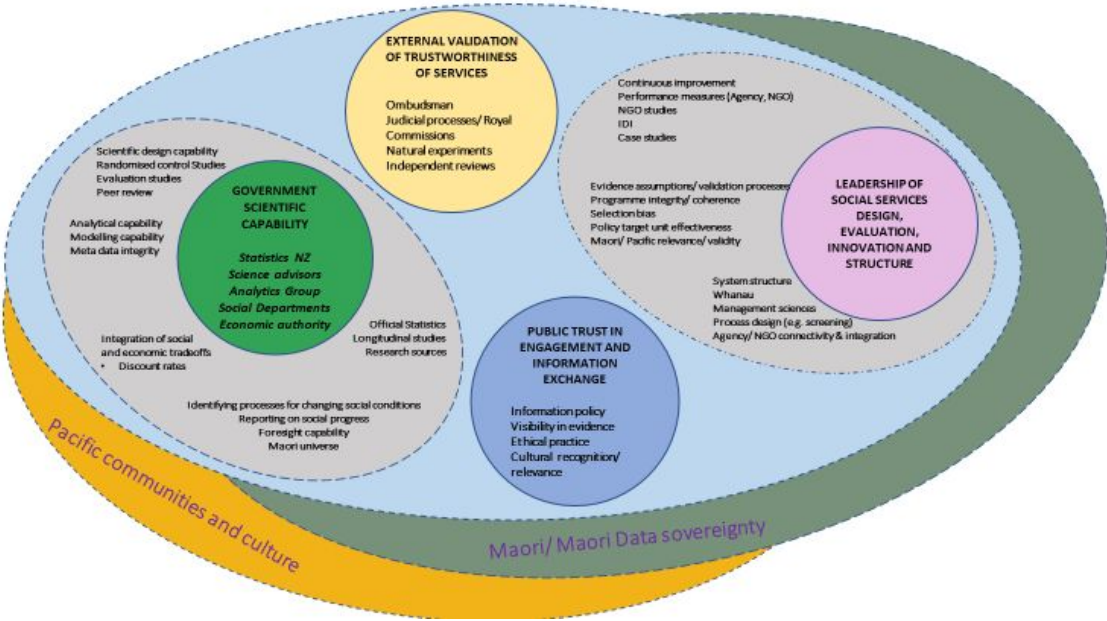
<sup>15</sup> NZ Treasury, 2010. Improving the Quality and Value of Policy Advice.

<sup>16</sup> See note 15 above

for building trust become weaker. Such attitudes can make using evidence for decisions in public policy very difficult. The reliability of social services policy and service delivery is dependent not only on the scope and of the underpinning evidence base but also the knowledge we can have about the practical limitations of measuring, summarising and modelling people. The potential evidence base for social services delivery ranges across performance measures, continuous improvement, evaluation studies and administrative records of past transactions, longitudinal and cross sectional statistical sources, forensic studies, meta analyses, as well as policy focused research, analysis and observations. All these sources of knowledge are an essential part of the evidence base for trustworthy social services.

Usually policy choices are informed by a tactical mix of science based evidence and analyses, individual observations, history, and insights. Qualitative or quantitative assessments of the quality of this evidence are less frequent. Choices that set the priorities for determining the scope, frequency and scale of specific evidence forms described above reflect the political mind set and institutional culture of those advising on and making choices. The sources of information that can originate from research from observational studies include official statistics, longitudinal studies, forensic analyses, administrative records, randomised control trials and continuous improvement. These sources enable modelling and analyses that can test the fit of policy to future populations, ethnic or regional communities, different cohorts and age groups. Political mind-sets and institutional cultures limit or expand aspirations, place significance on anecdote, are influenced by political and institutional ideologies and protect political and institutional identity. All shape attitudes to scientific evidence and its use, and how issues are framed.

Chart: The Spectrum of Evidence Sources for Ensuring Trustworthiness



The key fundamentals of the trustworthiness in the evidence behind social services delivery and practice are presented in the chart above, and include:

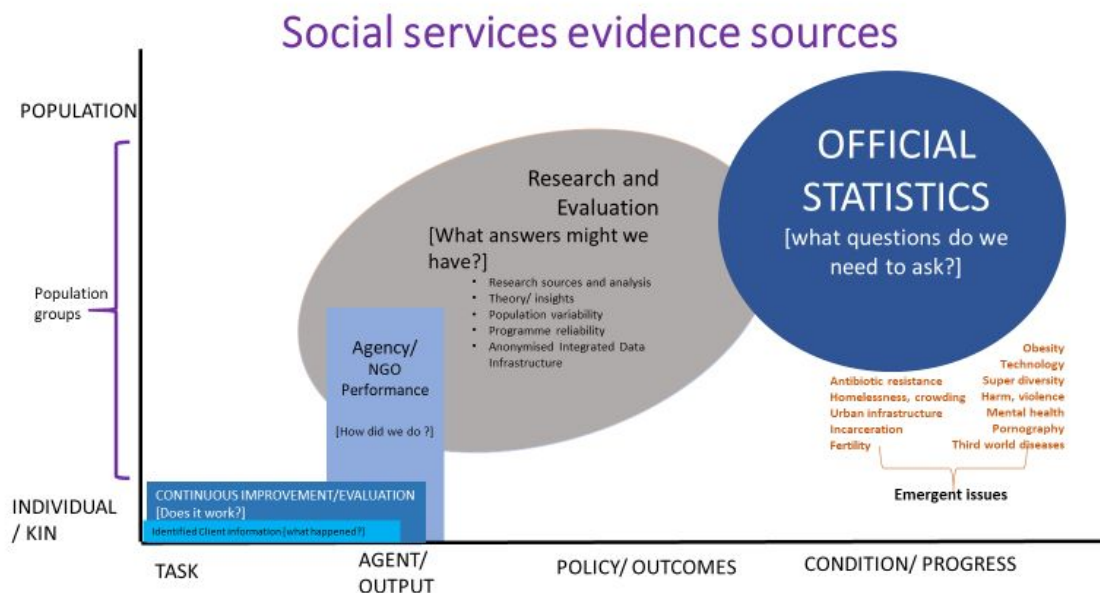
- the transparency of the political mind-set and institutional cultures in decision-making, that comes from operational and external processes of validation

- public trust in the gathering and exchange of information
- the relevance of the scope of evidence available
- the quality of evidence, including recognition and consideration of the diversity and variability of the human condition
- the integrability of the sources of evidence
- the analytical capability
- the design and oversight of operational practice
- transparency about the limits to the evidence
- ongoing evaluation of the practical application of policies.

## Scientific capability

Relevant and reliable evidence is an essential foundation for the analysis, selection, management, monitoring and evolution of social policy. The concept of a policy life cycle recognises that policy involves managing uncertainties that can often be taken account of by scientific methods of producing and analysis of evidence.

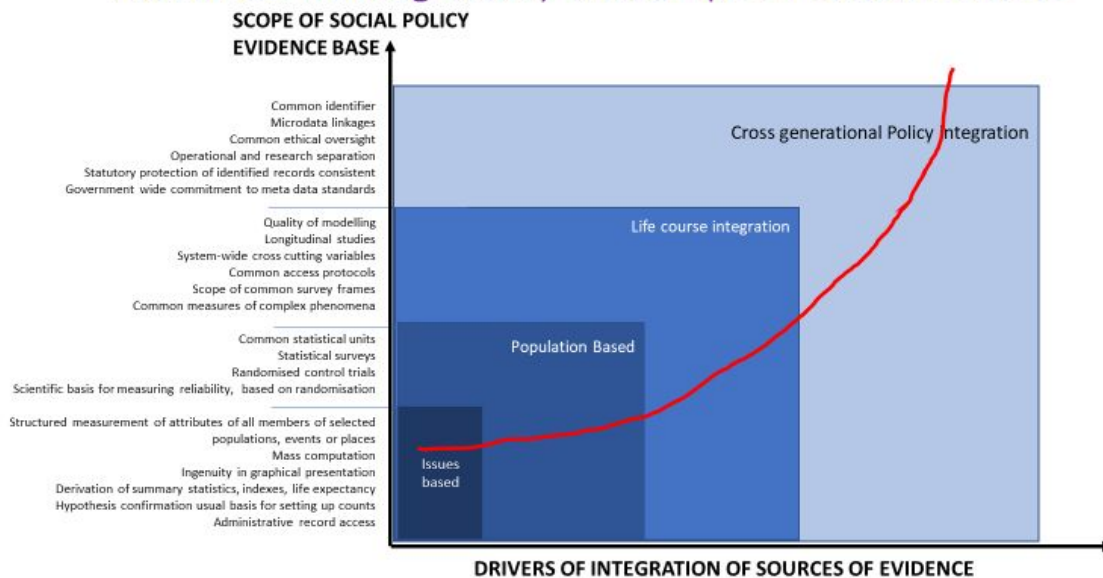
A systemic approach to increasing the scope and reliability of evidence would now include adding new sources of evidence, increasing the integrability of evidence sources, increasing the quality and connectedness of sources of evidence, expanding the range of analytical methods available, and widening the access to established sources and analysis. The chart below indicates the comparative place in the evidence system of well-known forms of information. The evidence forms each influence trustworthiness and quality in different ways; by the political and institutional context, the degree of relevance of the policy to the concern, practical issues of bias and variability, and the limitations of methodology.



## The significance of integrability on the scope of the evidence base

The scope of the evidence base for social policy and services will grow with the managed connectedness among sources of evidence that differ over time, place or culture, and the commonality between different attributes, characteristics and experiences in how they are measured. The degree of integrability is dependent on the commitment to a set of key cross cutting variables in all sources, which will be simplified where common identifiers exist. The most basic meta-data standards include common classifications and definitions, common delineation of key statistical units including family, whānau, cultural identity and location. This will also involve common population frames which result from national registers of population or addresses. The chart below demonstrates the exponential growth in the evidence base from a strong focus on different levels of integration.

### Advances in integrability and scope of evidence base



There has not previously been a strong commitment to the integration of government records across all sectors, although within education, health, taxation and welfare sectors the commitment had been stronger than elsewhere. The leadership of Statistics New Zealand of the IDI has signaled a significant transformation in the management of government records. The limited integrability of records managed by previous agency managers will reduce the immediate benefits through the lesser quality of the integrated records. Whenever the adoption of sector wide meta-data standards has become universal, the management of quality will be much simplified.

The external reviews of policy and analytical capability referred to earlier point to a need for a ground-breaking change in the valuing, gathering, managing and using evidence, and driving integration practices. Piecemeal advancement of strategies to increase the scope, quality and use of evidence in policy and service delivery may not advance the integration of the evidence we now have, either in anonymised forms for policy development and analysis, or in identified forms for connecting operations where there is benefit from doing so. Piecemeal strategies risk not identifying or recognising the benefits of system wide elements. They also risk fragmenting what must necessarily be broad based

approaches to deal with issues of privacy, trust and respect of citizens that must underpin the availability, integrability and quality of evidence.

As noted by Onora O’Neill, accountability and transparency can help build trust, but there are situations where that is not enough. Ultimately, the test for citizens is to do the right thing for them, and that requires setting expectations, recognising the variability and diversity of citizens, and the limits to the policies, services and rules that citizens face. Poor application of sanctions and penalties can destroy trust not only of individuals affected, but whole whānau and communities, for many years, and of other similar activities.

## Information forms and the evidence base

---

### Official statistics and the questions we need to ask

Official statistics play a vital part in reporting on the condition and progress of society, the economy and environment. They help citizens identify the questions that need to be asked to hold governments to account. Governments determine the scope of official statistics through funding decisions. Citizens have few means of holding government to account in this, but they can be highly effective. For example, official statistical surveys of disability were initiated only after the disability community threatened to boycott the 1996 population census, while the time use surveys resulted from a coalition forming agreement in 1996, after consecutive cabinet committees had earlier rejected the need for such information.

Official statistics are most important in pointing to the questions that we need to ask about our society, as they report on the absolute and comparative condition of groups, and the progress of groups and the community, making visible intergenerational comparisons and cohort differences. Official statistics are developed in full knowledge of the limitations of measurement, and most usually are associated with both qualitative and quantitative assessments of quality. There are wide ranging standards and frameworks for official statistics, ranging from the System of National Accounts to the Sustainable Development Goals, and many international forums where common definitions, classifications and practices are agreed on. Official statistics are generally stronger than other forms of evidence in how they justify trust through transparency in methods and openness about quality.

### A place for evaluation

Carol Weiss<sup>17</sup> noted that evaluation must compete for attention but also can be suppressed or result in watering down of evaluation conclusions because it doesn't fit the political process.

*“Evaluation research is a rational enterprise...But evaluation is a rational enterprise that takes place in a political context. Political considerations intrude in three major ways, and the evaluator who fails to recognise their presence is in for a series of shocks and frustrations:*

1. *“First, the policies and programs with which evaluation deals are the creatures of political decisions. They were proposed, defined, debated, enacted, and funded through*

---

<sup>17</sup> Carol H. Weiss: Where Politics and Evaluation Research Meet



*political processes, and in implementation they remain subject to pressures-both supportive and hostile-that arise out of the play of politics.*

2. *“Second, because evaluation is undertaken in order to feed into decision-making, its reports enter the political arena. There evaluative evidence of program outcomes should compete for attention with other factors that carry weight in the political process.*
3. *“Third, and perhaps least recognised, evaluation itself has a political stance. By its very nature, it makes implicit political statements about such issues as the problematic nature of some programs and the unchallengeable of others, the legitimacy of program goals and program strategies, the utility of strategies of incremental reform, and even the appropriate role of the social scientist in policy and program formation.*

*“Knowing that political constraints and resistances exist is not a reason for abandoning evaluation research; rather it is a precondition for usable evaluation research.”*

Ironically, just as evidence based information is reduced, gaps in knowledge will be replaced by anecdote, ideology, attitudes, simplified framing and limited aspirations. These may well undermine and delay recognition of the increased uncertainty that will affect the policy initiatives that they inform, and consequent operational performance.

Michael Cullen<sup>18</sup> pinpointed how bias occurs in deciding what evidence to gather, when he said of the 17th century statistician William Petty that;

*“...he was the first expression of his type to be found repeatedly in the history of social statistics: the reformer who saw the collection of facts as indispensable preliminary to practical and effective reform. The facts he chose to collect, as with later statisticians, were designed to demonstrate the necessity and desirability solely of those reforms which he desired.”*

### **Integrated Data Infrastructure**

The IDI provides rich opportunities to see the past transition pathways of targeted groups, and identify where there have been concentrations of people with experiences that could be better supported with this new knowledge. The information in the IDI monitors citizens through the lens of the state. The use of a confidential universal identifier within a protected environment has enabled the IDI to achieve a high level of integration across a wide of government sectors and traditional sources including the census of population. This integration over time has built an evidence base that can examine individual transitions, concentrations of attributes and characteristics, and potentially observe causality. There are limitations in the IDI in its early stages as most of the information obtained from contributing departments could not conform to public sector wide standards as there were none. Not all information recorded in administrative records has been managed consistently, as there is not yet the benefit of comprehensive and effective public sector wide meta-data management. There is a wide range of forms of evidence which taken together enable limitations in any one source to be mitigated. This leads to a need to balance deep expertise methods with strong sector specific knowledge, and to develop cross-sectoral approaches. The statistical properties of the IDI need to be determined at each application, as they are not determined by design. Randomisation and representativeness cannot be assumed, nor are

---

<sup>18</sup> Michael J Cullen (1975), *The Statistical Movement in Early Victorian Britain*

the variables gathered by the state necessarily those that have conceptual relevance to social science concepts.

Conceptually, the IDI is not about observing the actions of the state through the lens of the citizen. The experiences recorded by the state are specific to the statutory obligations placed on each evidence source, and may not reflect the full experiences of citizens with the agency. Not all citizens are either included or comprehensively observed in the IDI. All evidence sources of populations have some practical limitations, and some conceptual ones. The lives of citizens are much more variable than can ever be captured by the information gathered in research models or in administrative data collections. Consumer perspectives and their full experiences need to come from other sources.

The IDI is enriching the evidence base we have about how New Zealand's social services have jointly connected with people in the past. There is much we should get from its use, and it is important to recognise that these administrative records are able to provide a rich retrospective view on the past performance of agencies, and can inform improvements that have not been identified up to now. We can get much better knowledge about what didn't happen when it should have. The IDI could also be used to identify significant cross-agency transactions that may be best managed by extending the responsibilities of agencies.

The IDI has expanded the information with which we might estimate the likelihood of further engagement with social services by groups of individuals through linking records of individuals and facilitating analysis of the information of their experiences held within government agency records.

Information technology, particularly through the IDI, has added to the volume of evidence that can now be drawn on. There is a huge potential benefit from the effective management of the integration of disparate data sources. Longitudinal studies can bring an understanding of causality and rich insights about how behaviour and conditions might be influenced by interventions at a population, group or individual level.

### **Progressing towards a de facto population register**

Much of the driving force for the integration of information sources has come in the past from official statisticians, particularly driven by the needs of rich analytical frameworks including national accounts and demographic data systems, but also from the expanding focus on small area statistics. More recently, as approaches to targeting have become more specifically focused on individuals, then personal identifiers have become vital elements that ensure the veracity of the information gathered about individuals. While policy analysis and modelling need micro data about a sample of individuals that contains a relevant, rich set of variables, personal identifiers are not needed. Where services that are being delivered to a set of individuals by a mix of service providers, then sharing identified information becomes essential. Where that identified information is managed centrally, then de facto, there is a population register of service recipients which needs to operate with a high degree of trust and protection from misuse. Collective impact services need such a capability, not only because of the high trust needed between service providers and recipients, but among providers to have the necessary trusted availability of information, in real-time. The integration of service delivery requires information integration at a different timeframe, and certainty of exact personal identification that is less flexible than when evidence is used for policy analysis. The connectedness needed by citizens can be inferred

from the IDI, and this knowledge then used to identify where flexible organisation boundaries (e.g. social workers engaged by a housing agency) could be put in place.

The political acceptability of a formal national identifier has historically troubled New Zealanders (and Australians) yet the mobility of people not only across households and families is now such that location does not a strong determinant of identification for many. This reduces the effectiveness of place based services where long term follow-through and connection with several types of service providers exists.

## Modelling

Models enable the evidence contained in information sources to be distilled and summarised in a form that enables this comparison over time, across populations, groups and places, or applied to other information sources that share attributes in common. The strength of any associations derived in such models can be measured, and they underpin the quality of inferences that can be made by the model. For testing whether causality may be inferred, the information source usually needs to be designed for this. The information used to generate models is usually incomplete, not produced by a fully random process, nor is representativeness assured, while what is measured may only approximate or be a proxy for what is being tested. The validity of models depends on the strength of the assumptions needed to generate sufficient belief that the requirements of randomization, representativeness and realism will hold. Those assumptions may reflect political beliefs, theoretical concepts or ideological positions, or some methodological practice to induce elements of randomization. Because models shape policy, a lack of transparency in model assumptions can lead to expectations of objectivity in methodology that cannot be realised. Many model forms are well established and subject to methodological debate and transparency about limitations. For example, in population projection models, alternative approaches based on historical data of cohorts or period estimates are well understood. Similarly, there are alternative models in seasonal adjustment and for consumption functions. Measures of poverty can take several forms.

Predictive modeling is a term used currently to describe a form of modelling based on identifying individuals who will become the targets of services tailored to change their long-term prospects. Applying the results of such analyses is conditional on assumptions about commonality of characteristics across generations, the capacity of information recorded by the state to adequately describe the influences on prospective behaviours, and similarity of administrative practice over time. This limits to an unknown and unknowable extent the applicability of model parameters, predictions, rules and estimates of the likelihood of conditions and attributes. Neither operational rules nor analytical models generally take this uncertainty and potential for bias into account. The veracity of predictive modelling is limited by the ecological fallacy, where the attributes of a group are erroneously assumed to be applicable to individuals in the group instead of just the group collectively. McIntyre and Ellaway<sup>19</sup> note that the 'ecological fallacy' involves inferring individual-level relationships from relationships observed at an aggregate level.

In discussing the limitations to the accuracy of modelling from partial populations using historical administrative records, Keddell<sup>20</sup> gives examples of the causes and potential scale of type 1 and type 2 errors and their implications for children when using administrative records of past children to identify

---

<sup>19</sup> S McIntyre and A Ellaway, Chapter 14, Ecological approaches: Rediscovering the Role of the Physical and Social Environment. *Social Epidemiology* Berkman and Kawachi (2000)

<sup>20</sup> Keddell, E. (2016). Substantiation decision-making and risk prediction in child protection systems. *Policy Quarterly*, 12(2), 46-56.

screening characteristics that predict individual future need for child protection services. In the study being critiqued, “While 12 different algorithms were tested, the most successful one concluded that the three most significant predictors of substantiation were: length of time spent on a benefit; contact with Child, Youth and Family [CYF, now Oranga Tamariki] as a child; and the substantiation of other children in the family”. Given that Māori and Pacific families generally have larger family sizes, such a prediction model is going to have a bias towards assigning higher risk identities to Māori and Pacific children, all other things being equal. Administrative records do not contain information recognised conceptually as potential predictors.

## **Leadership of social services structure and design**

---

### **Opportunities from system-wide leadership in the evidence base**

The nature and structure of the social services system itself might change if we were alert to the potentially knowable limitations on the quality and scope of evidence that underpin social policy and service delivery. Providing knowledge about the quality of information that informs the delivery of social services should uplift the level of trust in the services by consumers. Expanding the autonomy of front line might improve the value consumers get from services with untested reliability. There are limits to the exchange of information between organisations, and we could consider organising groups of organisations around the quality and scale of information flows, and the limits to the capacity to exchange information. The collective impact approach, of which Whānau Ora is an example, is where the most important connections between services and the service recipient are embedded in the institutional arrangements within and between the involved organisations. Knowledge of the lack of precision in measures and models should inform the relative priority we give to interventions, especially if we were able to relate their cost to measures of effectiveness and efficacy. The balance between targeted and universal services might subject to more effective challenge if we knew how the comparative lack of precision in the evidence base affected operational reliability.

There is now a major opportunity to improve the quality of social services through the broadening of the evidence base now available, the increased integrability of new and existing evidence sources, and enriched analysis. Major advances in the use of evidence will necessitate broadening the application of well-established methods including continuous improvement and management sciences such as operations research. Much more could still be done with simple measures to monitor services. Performance measures of departments have a powerful influence on agency priorities and risk assessment and management, and can also influence culture, relationships and aspirations for the public that they deal with. The narrowing of agency responsibilities that was central to the public-sector reforms of the late 1980s has resulted in rigid boundaries between organisations operating independently on the social services sector. This is not unlike how the British leadership of the desert war operated before Montgomery took over<sup>21</sup>. One expert<sup>22</sup> estimated from his extensive amateur athletics experience that while the chance of a relay runner dropping a baton is a about 0.1 percent when running their leg, dropping a baton is some 40 times more likely at the point of passing the baton to the next runner. Transferring information may be less fraught than a high-speed baton transfer, but the risks are still likely to be significant. This has limited the adaptability of agencies to the huge shifts in

---

<sup>21</sup> Willoughby Norrie “... *the cowpat theory of war, where your troop dispositions are widely spread like cowpats in a paddock, one here, one there*” - as quoted in Kippenberger, by Glyn Harper

<sup>22</sup> Personal discussion with Dr Frank Nolan, statistician and athlete.

the complexity of demands that a diverse population has brought, over a period when not only their current but future responsibilities were narrowed. Consultation with people who use a mix of services would identify areas where increasing at the margin the scope of an agencies activity would improve service quality. For example, in the past, the housing agency has employed field workers who could resolve issues involving renters, without involvement other agencies. The side-effect of autonomous agencies is well illustrated in the David Low cartoon from 1932.



'PHEW! THAT'S A NASTY LEAK. THANK GOODNESS IT'S NOT AT OUR END OF THE BOAT' (1932)

The currently evolving organisation and structure of the social services system itself could be informed and challenged by the expansion of the evidence base. The quality of the components of the evidence base ought to be taken in to account, as should the limitations on its effective transfer. Such awareness could inform questions about:

- the limits to connectedness of organisations
- defining the scope of activities of delivery organisations
- rethinking the boundary between universal and targeted services
- the relationship between the autonomy of front line staff and the quality of evidence
- how interventions are prioritised
- the necessity of continuous improvement
- the importance of quality frameworks.

Any such rethinking will need to recognise that the judicial system is a distinct branch of government, with a distinct constitutional role in the oversight of the legality of the activities of citizens and government. The judiciary provides assurance that the government of the day and the agents of the

state can be held to account for their actions. This separation brings constitutional bounds to the integration of the justice sector, as recently commented on by the Chief Justice<sup>23</sup>.

## Evidence that services work – continuous improvement

W Edwards Deming's universal principles to managers for transforming business effectiveness are embedded in the many shades of approaches labelled Total Quality Management. Deming articulated fourteen principles, and some examples of where their adoption would most likely have changed outcomes are given below.

1. *Create constancy of purpose toward improvement of product and service, with the aim to become competitive, to stay in business and to provide jobs.*
  - **CYF lack of monitoring of exits of young persons at age 17 years**
  - **Waitangi tribunal case on recidivism performance measures**
  
7. *Institute leadership (see point 12 and chapter 8 of Out of the Crisis). The aim of supervision should be to help people and machines and gadgets do a better job. Supervision of management is in need of overhaul, as well as supervision of production workers.*
  - **Ombudsman<sup>24</sup> inquiry into prisoners at risk of suicide at Corrections**
  - **MSD monitoring of lost records<sup>25</sup>**
  
9. *Break down barriers between departments. People in research, design, sales, and production must work as a team, in order to foresee problems of production and usage that may be encountered with the product or service*
  - **Havelock North town water failures**
  
12. *Remove barriers that rob people in management and in engineering of their right to pride of workmanship. This means, inter alia, abolishment of the annual or merit rating and of management by objectives (See chapter 3 of Out of the Crisis).*
  - **Housing New Zealand (HNZ) methamphetamine test being not fit for purpose**
  - **Perceived constraints in the Privacy Act 1993 can limit information sharing to the disadvantage of citizens**

Continuous improvement practices have been long embedded in the work of the public sector in Inland Revenue, New Zealand Customs Service and Stats NZ, but are not prominent in the social services sector. Examples of poor application of continuous improvement are given below:

- Free access to pure water is a vital part of our health system. When water no longer deserves the trust we place in it, then its use should be prevented as soon as it is known to be unsafe. In Havelock

---

<sup>23</sup> Dame Sian Elias, Chief Justice of New Zealand: Address to the Criminal Bar Association Conference, August 2017 'Managing Criminal Justice'

<sup>24</sup> Ombudsman 1 March 2017: "Care and management for prisoners considered to be at risk of suicide and self-harm: observations and findings from OPCAT inspectors". In April 2016, we requested the following information on tie-down beds from Corrections:

1. Which sites have tie-down beds?
2. Which sites have used the tie-down beds between 1 April 2013 and 12 April 2016?
3. On how many occasions have they been used?
4. What was the duration of each tie-down episode?
5. How many prisoners have been secured on tie-down beds during this period?

The Department informed us there is no central recording system for documenting tie-down bed use and that individual prisons do not record the information in logbooks

<sup>25</sup> OI Request: "Can I have a summary that shows the number of times a beneficiary has to return to MSD because their documents or any part of them has been lost by MSD when servicing the request. Information for the last five years if available please." Response 30/3/2017 from MSD "The information you have requested is held in notes on individual case files. In order to provide you with this information Ministry staff would have to manually review thousands of files."

North drinking the town water became a cause of illness rather than part of the cure. People continued using polluted water as though it was assumed to be safe. It appears that water was tested regularly at 48-hour intervals, regardless of changes to the context in which it was gathered, and despite a history of adverse events. Constancy of purpose did not imply constant practice when the context could vary significantly.

- The severity of response generated by an administrative process needs to be accompanied by the appropriate certainty of guilt. After two Court cases to evict tenants were dismissed, HNZ admitted that their test to measure the presence of methamphetamine was not fit for purpose. The seriousness of methamphetamine addiction and the harms it causes is well recognised. The cost to individuals associated with methamphetamine is high – being blacklisted by HNZ, massive clean-up bills and possible oversight of children by Oranga Tamariki. The discovery that the prevalence of methamphetamine at levels above the HNZ thresholds was widespread (even on banknotes) has brought these tests into disrepute, not only by the Courts. This again should be cause for reflection.

All these very topical examples point to the potential of weak or erroneous evaluation processes to obscure rather than pinpoint whether what is being done is both worthwhile, and the right thing to do. The outcome of poor or no evaluation can in fact lead to harm. These examples generally highlight an absence of evidence about what makes a good experience for citizens, and highlight the harm that can result from insufficient commitment to quality management, and a narrowed framing of issues

## Māori

---

The proper recognition of Māori in social services requires not only having Māori recognised as a distinct analytical entity in information sources, but acknowledging their characteristics in the design of those sources. Determining the priority given to statistical sources is much deeper than establishing the rich distinct source such as Te Kupenga, but also requires acknowledging the different importance of statistical sources such as disability and time use to Māori. Performance measures can obscure or ignore the distinct interests of Māori, as exemplified by the Treaty of Waitangi Tribunal judgment<sup>26</sup> criticizing the performance measures of the Department of Corrections.

Screening methods can be influenced by differences in the aspirations for different groups that are implicit in policy, or in the expectations of those that deliver social services programmes. Māori have had different experiences from the same system because of this, and sometimes aspirations are constrained by those who are themselves consumers of the social services system. Evaluation can identify where this occurs.

We have long known that the Māori population has different demographic characteristics, is spread differently around New Zealand and has different family and community structures through whānau and hapū. Māori still experience outcomes in health, education and employment that are outside the norm of those systems that deliver services. Experience has been to apply solutions that placed little importance on long term remedies relevant to the position of Māori or their place in the determination and application of services. The practices of service delivery and evaluation need to be aligned to the

---

<sup>26</sup> Waitangi Tribunal and Maori (April 2017) “In 2014, the Department [of Corrections] declined an Official Information Act request from Mr. Hemopo seeking to understand how the claimed reductions in reoffending affected Māori in particular. It declined the request on the basis that as ‘The Department does not calculate Better Public Services targets reductions in re-offending results separately by ethnicity...the documents alleged to contain the information requested does not exist’.”

characteristics of the populations involved, to avoid systemic biases. Many of our approaches reinforce deficits for Māori and ignore the strengths and opportunities that exist within whānau and hapū to create change for themselves.

We know little about whānau as a multi-faceted source of wellbeing. Whānau span many forms of functionality, and what is an enduring element is whakapapa. Without the protection of whakapapa, the enduring cross-generational connections to others and to the land are lost. Whānau can be regarded as capital that is embedded by whakapapa. The depth of whanaungatanga is generally ignored in policy. Whānau and whakapapa are about people and place, and embrace matters of either, including urban planning and housing. Communal style living still has a place for some whānau, which can complicate aspects of contemporary life. Whānau members are inherently mobile, which is more difficult for later generations. These characteristics of whānau are generally not detectable from long standing sources including the census. There is a need for to draw on Te Kupenga more strongly in understanding whānau. Not accounting for whakapapa is a matter of concern for other ethnicities. Across a multiplicity of cultures, there is likely to be a need for more than one lens based on whakapapa to understand matters of social, cultural, economic and environmental concern, and matters of justice.

## **The place of uncertainty in informing social services policy, delivery and structure**

---

### **The many natural sources of uncertainty**

Social services can aim to influence characteristics and conditions, or just ameliorate their current and future consequences. Mechanisms that are based on the causality between practice and outcome usually result from scientifically designed research studies, but service delivery necessitates applying their results to population groups different from those for whom the research results were validated. Cultural differences, gender, age or other effects bring a need for the validity of the results to be separately justified for the target population of the times. Understanding the diversity and variability of the human condition, and how to take account of it may vary over time, and differ with each generation and among cultures. The observations of people and processes that form evidence in the social services are rarely precise, or complete in who they represent, even when they contain all who are part of some service or condition. Knowing what we can about the quality of evidence underpins interpretation of measures and models, and can shape the nature of service delivery when its form is based on this evidence.

We often do not have evidence that is specifically focused on the concern at hand, so accumulating evidence about similar processes can help approximate the information that would have been obtained were direct observation possible. Sometimes our tools are not up to the job. For example, comparing different types of benefits over very long time-periods requires evaluation tools that can cope with valuation complexities, risk measurement and inflation, as well as recognising indirect benefits and costs.

We can assess the reliability of service delivery processes based on scientific approaches to measurement, modelling and analysis unlike those services whose information base has been limited the political mind-set or institutional cultures. Evidence can have multiple origins, and most evidence was not designed for many of the purposes to which it is put. It is essential that there are qualitative or



quantitative measures of uncertainty, so that policy and service delivery can be developed with some knowledge of the limits to the ability to judge whether evidence and analyses are fit for purpose (depending on the question). There are well established frameworks for assessing quality in many fields, such as official statistics, which are underpinned by elements that are fundamental to quality assessment in any process. Without some qualitative or quantitative indicators of the uncertainty of the available evidence and programme risks, it is quite difficult to determine whether any visible adverse events represent system failure, or are an outcome with known likelihood from a process with known variability. Models shape policy, and the assumptions that lead to their acceptance need to be validated, both generally and in the context of their use.

Inadequacy in measurement and assessment can bring unintended consequences including unnecessary costs to government and citizens, personal risk and inappropriate penalty. Any social service process involves risks about reliability which will not be controllable without a strong commitment to management sciences. In the social services, agency performance measures generally are not designed to signal changes in the reliability of services received by citizens as consumers, or changes to the cost to consumers of engagement. Management sciences that range from operations research to continuous improvement can bring a rigorous approach to improving services regardless of the integrity of the evidence base for the underpinning policy, and bring untapped opportunities to exploit the currently expanding administrative data sources.

## **Pacific**

The Pacific communities have a different population dynamic than New Zealanders generally, one which may be more comparable to Māori than any other community. These generate preferences about child development and youth that get submerged in policies assessed by national targets. Although they are young communities, the rate of growth of the elderly in these communities is faster than that nationally. Pacific communities are rarely distinguished in the analytical base of policy, and their distinct community and extended family characteristics are generally ignored in service delivery.

## **An indicative measure of selection bias**

The potential scale of selection errors can be seen from Stats NZ information about disabilities, which is obtained by two comparable and well-regarded statistical sources from the same people. The population census and the disability surveys whose survey frame is drawn from census responses provide a natural experiment for measuring selection bias. The trigger question in the population census is always very simple, while the survey seeks a more exacting classification.

There have been several forms of questions about disability in the censuses since 1996, including 2001, 2006 and 2013. The responses are used to select a sample of people for more comprehensive and exacting enquiry about their disability status and aspects of their lives. The survey questions are contained in the Stats NZ paper<sup>27</sup> referenced below. The detailed methodology results are presented in the appendix.

---

<sup>27</sup> Statistics New Zealand (August 2015) Measuring disability in New Zealand: Current status and issues

**Figure: Disability questions in the 2013 Census of Population**

**16** Mark as many spaces as you need to answer this question.  
Does a health problem or a condition you have (lasting 6 months or more) cause you difficulty with, or stop you from:

- seeing, even when wearing glasses or contact lenses
- hearing, even when using a hearing aid
- walking, lifting or bending
- using your hands to hold, grasp or use objects
- learning, concentrating or remembering
- communicating, mixing with others or socialising

or  no difficulty with any of these

**17** Do you have a long-term disability (lasting 6 months or more) that stops you from doing everyday things other people can do?

- yes
- no

These two enquiries provide false positives or type 1 errors of 28 percent, and false negatives or type 2 errors of 15 percent from the population census screening, compared to the post censal survey.

If the error rates of screening processes were generally of this order of magnitude, then operational practices based on the use of linked historical administrative records obtained from a multiplicity of agencies would have limitations in the automated selection processes now being developed. Each such process must have some form of independent validation to indicate the limits to their precision. For other forms of screening, any shift in how responses are classified will need to be tested and controlled for, given their importance in key areas in determining need for referral elsewhere. An example is the move to accelerate the establishment of call centres accelerate. As they become more prevalent in areas of family violence, harm and protection, then it should be possible to identify the net gain from offsetting a possible increase in service variability with the greater benefits from increasing the range of people who can conveniently, quickly, and safely access a service that could offer more services. For young people in New Zealand, texting is highly preferred (average texts per month of 230 compared to 25 mobile calls<sup>28</sup>)

## Conclusion

---

Science can bring evidence that provides new options for programmes that can ameliorate, reduce or prevent conditions that reduce the quality of life for people during their life course. Services need to function impartially regardless of the quality of the policy setting that led to their being put in place. Processes to continually improve the outcomes for service recipients need to have full transparency about the limitations of the underlying policy, and the programme instruments. Without this, citizens can play a high price in drawing on services of uncertain effectiveness, and in some cases, can be placed at high risk. The examples given serve to illustrate this. In all forms of evidence there are quality risks that need to be actively managed. The way ahead involves making explicit judgements about quality and looking to forms of testing and evaluation that are relevant to the policy, services and populations served. The full spectrum of evidence available in social services is often ignored, despite its importance to policy development and critical necessity for service delivery as well as the structure of social services.

---

<sup>28</sup> Haxell 2015 On becoming textually active at Youthline, New Zealand

## Key themes

- Science, policy and practice are not naturally well connected – political sentiment and institutional cultures can dominate policy and operational choices, and the selection of models which justify them.
- Society and people are diverse, hard to classify, making screening and targeting more fraught than we recognise (Statistics NZ census study).
- Integrating of social services is limited by complexity.
- One essential underpinning of trustworthiness is continuous improvement, which is especially important in the face of the limited evidence base for much policy determination.
- Having widely trusted contextual information can play a significant part in assurance of the trustworthiness of policies and their prospective value, dependent on how choices are made.
- Practices such as predictive modelling need to meet the same tests as other trusted sources of evidence, in terms of methodological transparency, visibility of assumptions and their validation, and ongoing monitoring through continuous improvement and periodic review. The ecological fallacy limits the validity of some approaches.
- Where methods of measurement, modelling and analysis affect the lives of groups of people, by determining the delivery of services at a personal level, then their quality needs to be made transparent by both qualitative and quantitative assessments.
- Not only the content but the quality of the evidence base may influence the path of the evolution of the social services system.
- There are strong quality frameworks that are used by statisticians in many fields that need to be adopted by those involved in measurement and modelling to make transparent the limits to the scientific integrity of the methods and information sources in the context of their application.
- Any quality framework must recognise what is needed to be relevant to Māori, and make transparent how the distinct characteristics of other cultures are taken account of.
- Piecemeal strategies are unlikely to achieve all the benefits of increased data access and may slow how to deal with issues of privacy, trust and respect of citizens that must underpin the integrability of evidence. Ground-breaking change is needed in the valuing, gathering, managing, integrating and using evidence.
- The evidence we have about the connections that citizens need to make between organisations could determine organisational roles, staff autonomy and information exchanges, given the complexity and scale of those with multiple connections to social services delivery agencies.

## Wisdom on trustworthiness – the philosopher, statistician, poet and politician

<p><b>Onora O’Neill</b></p>	<p><i>“To be accountable is not merely to carry a range of tasks or obligations, for example to provide medical treatment to those in need, to make benefit payments to those entitled to them, or to keep proper accounts.</i></p> <p><i>“It is also to carry a further range of second-order tasks and obligations to provide an account of or evidence of the standard to which those primary tasks and obligations are discharged, typically to third parties, and often to prescribed third parties.”</i></p>
<p><b>W Edwards Deming</b></p>	<p><i>“Ninety percent plus of all problems in variation or defects are the result of the system rather than the individual.</i></p> <p><i>“Improve constantly and forever the system of production and service.”</i></p>
<p><b>Choruses from The Rock T.S. Eliot, 1934</b></p>	<p><i>I say: take no thought of the harvest, But only of proper sowing.</i></p>
<p><b>Prime Minister of France, L Jospin (opening International Statistical Institute session, Paris, 1989)</b></p>	<p><i>“The right to information has become one of the fundamental rights of the twentieth century citizen. In a society where information and the media play a considerable part, your [professional statisticians] action helps safeguard a fundamental human liberty...The working methods you use are complex, the data you deal with difficult to evaluate. An effort to explain [to the public] is necessary. This effort is required by democracy. <b>All citizens must be in a position where they can understand and assess the policies followed by governments.</b>”</i></p>

## APPENDIX: Opportunities and constraints in the evidence base for social policy and services

Shaping the evidence base in social services		Influences on the quality of social services policy and delivery		
Quality dimension	Quality elements	Political and institutional constraints	Practical limits to the evidence	Tools for managing the quality dimensions
<b>Political and institutional context</b>	<p>Scope of evidence base</p> <p>Integrability of evidence base</p> <p>Analytical capability</p> <p>Making research known (competence, reliability)</p> <p>Policy framework</p> <p>Knowledge frameworks</p> <p>Protocols, conventions, laws, roles and institutions</p> <p>Institutional practice and incentives</p> <p>Diffusion of knowledge</p>	<p>Poor information management practice</p> <p>Poor historical infrastructure</p> <p>Weak analytical capability</p> <p>Underuse of evidence base</p> <p>Privacy issues unresolved</p> <p>Family contribution undervalued</p> <p>NGO knowledge ignored</p> <p>Discount rate adaption slow</p> <p>Performance focus misdirected</p> <p>Cultural biases</p> <p>Human rights obligations put aside</p>	<p>Currency of concepts</p> <p>Gaming</p> <p>Untested use of Analytical reasoning weak</p> <p>symmetric loss functions</p> <p>Simplistic methods</p> <p>Gaming by agencies and citizens</p> <p>Cultural capital</p>	<p>Science advisors</p> <p>Transparency</p> <p>Standards of evidence</p> <p>Protocols for release of research</p> <p>Centres of excellence</p> <p>Sector standards for continuous improvement</p> <p>IDI</p> <p>Meta-data management</p> <p>Social licence</p> <p>Cultural capital</p>
<b>Relevance of framing of issues</b>	<p>Fit of measures to models</p> <p>Time relevance</p> <p>Comparison of monetary value of costs and benefits over long periods of time</p> <p>Emergent issues</p>	<p>Incomplete/partial description of characteristics</p> <p>Asymmetry of risk to individuals from type 1, type 2 errors</p>	<p>Use of proxy variables</p> <p>Static nature of models</p> <p>Averaging effects</p> <p>Discount rates</p>	<p>Analysis of natural experiments</p> <p>Peer review</p> <p>Cross disciplinary teams</p> <p>Cultural frameworks</p> <p>Cultural accountability mechanisms</p> <p>Performance measures</p>
<b>Accuracy - bias</b>	<p>Representativeness, randomisation</p> <p>Evidence limitations</p> <p>Selection bias</p> <p>Service barriers</p> <p>Incomplete/partial description of characteristics</p> <p>Response failures</p>	<p>Operation failures</p> <p>Averaging effects ignored</p> <p>Using summary measures instead of distributions</p>	<p>Willingness to report</p> <p>Inconsistent data definitions</p> <p>Non-representativeness of cases</p> <p>Partial or delayed responses</p>	<p>Transparency of methods</p> <p>Statistical expert review processes</p> <p>Knowledge accumulation initiatives</p> <p>Evaluation standards</p> <p>Peer review</p> <p>Standard error frameworks</p>
<b>Accuracy - variation</b>	<p>Variation untested/ unaccounted for</p> <p>Errors in variables</p> <p>Sampling error</p>	<p>Service variability</p>	<p>Variation untested/ unaccounted for</p> <p>Measurement errors</p> <p>Variability and diversity of people</p>	
<b>Limitations of methodology</b>	<p>Modelling misspecification</p> <p>Over-simplistic methods</p> <p>Confounding variables</p> <p>Static nature of models</p> <p>Lack of expertise/ experience</p>	<p>Ecological fallacy</p> <p>Confounding of system effects with individual results</p> <p>Naive analysis (Meadows and cot deaths<sup>29</sup>)</p>	<p>Population mobility</p>	<p>International peer review</p> <p>Centres of expertise</p> <p>Transparency of methods</p> <p>Leadership of analytical communities</p> <p>Expert advisors</p> <p>Management sciences</p>
<b>Feedback from evaluation</b>	<p>Accountability and oversight</p> <p>Continuous improvement</p>	<p>Investment in evaluation Natural experiment learning limited</p> <p>Weak commitment to evaluation and continuous improvement</p>		<p>Accountability processes</p> <p>Evaluation frameworks</p> <p>Consumer centric monitoring</p>

<sup>29</sup> Philip Dawid, (2008). Statistics and the Law: In Evidence, ed A Bell et alia