USING EVIDENCE TO INFORM SOCIAL POLICY: THE ROLE OF CITIZEN-BASED ANALYTICS.

A DISCUSSION PAPER

Sir Peter Gluckman
19 June 2017

This paper is the first in an occasional series arising from a research partnership between the Government of New Zealand and the European Commission, which is exploring the opportunities and challenges related to investing for better social outcomes arising from progress in data analytics and improved understandings of relevant human sciences – what has been termed by the NZ Government, "social investment". During 2016 it was agreed that the Office of the Prime Minister of New Zealand's Chief Science Advisor would lead a project jointly with the European Political Strategy Unit and the Joint Research Centre of the European Commission to explore the challenges, opportunities and methodologies associated with the evolving use of big data in social policy domains. There are commonalities between this approach and work being carried out in the Joint Research Centre in the areas of fairness and resilience. Future papers in the series will explore both the conceptual and practical aspects of combining the growing understandings in biological, behavioural and social sciences with big data analytics to assist social policy development and evaluation.

EXECUTIVE SUMMARY

This paper considers both the considerable opportunities and the issues associated with using individual-level data and citizen-based analytics to inform social policy development and implementation. This growing application of big data analytics occurs within the context of new insights from biological, behavioural and social sciences that allow for greater understandings of how data can be better applied to assist in the evaluation both *ex ante* and *ex post* of social sector interventions. In New Zealand, this approach is being developed to underpin social policy development and implementation and has been termed 'Social Investment'. In Europe, this type of analysis aligns with ongoing work of the European Commission's Joint Research Centre in the areas of fairness and resilience.

The basic argument underpinning this general approach is that interventions at one stage of the life-course will have impacts later in the life-course, and often with a much broader range of outcomes than those initially targeted. While this general understanding is not new, the scientific, behavioural and social science that seeks to explain these relationships has advanced rapidly, allowing for a broader range of possible policy developments to be considered.

The emerging challenge for policy-makers is to better understand these complex relationships in a way that allows them to decide which interventions, delivered to which individuals, at what stage in their life-course, will do the most to boost resilience later in life - leaving individuals better able to cope with the inevitable stressors of life. Because there is an almost infinite range of possible social interventions, governments are placed in a position of having to choose between many possible options. Each has its advocates and the decisions are inevitably politically charged. Further, the evidence needed on which to decide to terminate an ineffective programme, or to enhance an effective one, is often absent.

Individual life-courses are affected by many influences including biological, family, social and environmental factors. Governments and social sector providers are increasingly able to collect and use knowledge related to a growing number of these factors. By combining such data, especially longitudinal and multi-domain data, in large databases under the appropriate guidelines and controls, it is possible to obtain group-level data that provides a greater understanding of how these factors interact, and of the potential for prevention, amelioration or remediation. This is the basis of 'citizen-based analytics'. The compelling scientific, ethical, economic and policy arguments for this approach are presented.

From both the State's and the individual's perspective, optimal social outcomes require integration and delivery of effective and efficient services. Citizen-based analytics must be supported and informed by scientific understandings to avoid a large number of potential interpretative errors. Whereas citizen-based analytics generally only need anonymised but

individually linked data, there are some circumstances in service delivery assessment and management where identifiable client level data may be. Therefore, it is important that citizen-based analytics are supported by appropriate data governance and safeguards, accountability and oversight that take into account and justify the distinct purposes for which data may be needed.

This paper discusses these distinct dimensions, including: the need for social licence and transparency around data use; the need for governance and data management structures; the need for clarity of system architecture; the importance of high quality data; and a number of related issues including the data use for service improvement. Citizens need clarity about the purposes for the collection, curation and use of their data. The paper discusses the considerable utility and also the limitations of citizen-based analytics, the broader implications of this approach, and the potential impact on the policy process and service delivery.

.

Accompanied by the appropriate use of empirical policy trials and client level information, there is considerable potential for progress in social service delivery by the evolving use of citizen-based analytics. However, to achieve this will require consistent and ongoing attention to transparency, social acceptance, strong data governance, a commitment to data hygiene, curation and quality, upskilling within the policy community, continued vigilance against poor data analysis and interpretation and algorithmic bias, as well as ongoing engagement of policy-makers, service providers, academics and the wider community. Citizen-based analytics can provide new tools for liberal democracies seeking to create better outcomes for their citizenry.

New Zealand's 'social investment' approach is highly innovative and at the cutting-edge of applying citizen-based analytics for social policy development. The quality of the analysis and its utility will grow as data science and social science are jointly applied, as databases become more complete and as various issues are addressed. This should in turn drive even more effective social policy development and service delivery. It is emphasised that the primary benefit of this approach is in terms of 'social good' returns, and that fiscal (savings) estimates are simply rough proxies of that subset of benefits that can be readily estimated

Introduction: The policy challenge

Arguably, one of the biggest challenges in public policy is effective decision-making regarding social sector interventions. All policy stakeholders ultimately desire optimally effective and efficient services. Clients want services that meet their needs, taxpayers don't want their money wasted, governments want to see better outcomes for the population to whom they are responsible; the moral, social and economic arguments align. But it is a complex and contentious area of policy making precisely because lives are complex and individuals and societies are complex and growing more complex. The consequential diverse views about priorities, cause, cost and benefit are all values-based and will vary enormously.

Typically in advanced liberal democracies, decision-making regarding social sector policy is the result of policy trade-offs between a combination of normative[1] positions, political ideology and electoral considerations, supported by limited empirical evidence. Often the scientific and evidential base is overpowered by these broader dimensions. Arguments of effect, cost, benefit and spill-over effects - both positive and negative - are not always informed by rigorous evidence. A firm evidence base is often lacking for many programmes that have been introduced and, in general, few programmes are subjected to any formal, independent evaluation. Further, there is a general inability to effectively evaluate whether a new programme will significantly complement, supplement or replace an existing programme. Even where programmes are piloted, issues of scalability and context may limit their broader utility and effectiveness. The nature of social sector programmes is that once in place, it can be very difficult to determine effectiveness and indeed stop those that are likely ineffective, a problem which is inevitably compounded by advocates of the particular policy and amplified by the interests of providers associated with the programme delivery.

The general approach to policy making in the social sector is based on developing four levels of understanding, each of which has challenges:

1. Clarity as to the goal(s) of any particular intervention
2. Understanding what segment(s) of the population should be targeted
3. Understanding what works for them in what context
4. How to deliver effective services (from the perspective of the client as well as the funder).

Social sector expenditure involves a mix of targeted and universal entitlements comprising remedial expenditures (e.g. justice, health care), maintenance expenditures (e.g. unemployment and disability benefits, pensions etc.) and investment expenditure (e.g. education, vaccinations etc.). Where the balance lies across these types of expenditure is at the heart of much of democratic debate and partisan positioning. Much of the debate is

---

[1] 'Normative' refers to judgments that are derived from established values, cultural norms or standards of behaviour rather than made on the basis of empirical evidence.

underpinned by differences in theories of social justice. Therefore while views as to where that balance lies vary markedly, the general goal is fundamentally about achieving a cohesive and fair society in which as many members as possible develop and sustain the social, cognitive and physical capacities to be productive, and where those who cannot – either transiently or permanently – are appropriately supported. It is also about ensuring that supportive social (e.g. family) and physical (e.g. housing) environments are created. Further, it is about ensuring that social cohesion is protected through fair and appropriate policing and judicial decisions.

This is the context of ever-evolving social systems. With increasing modernisation, society has increased in its complexity - for example through changes in demography due to altered family structures, reproductive behaviour and life-span, through changed social attitudes and through altered labour practices and migration. The impact of digital and other technologies on how people live their social lives cannot be underestimated[2]. The result for public policy has been a progressive increase in targeted as opposed to universal entitlements over a period when society has become increasingly diverse.

These various dimensions create many contextual and individual factors that affect how any individual's life in any society evolves. Social policy makers thus have the challenge of identifying those factors that will enhance the lives of and opportunities for members of a jurisdiction and protect those who are vulnerable, assist those who are in distress, protect individuals from antisocial events, punish and rehabilitate those who commit anti-social acts and ensure that society as a whole continues to evolve in a positive manner.

All the while, it is necessary to acknowledge that society meets these goals through resource transfers – thus creating the need for transparency and a social contract between taxpayer and beneficiary (noting that often this is the same person or will be at different times and stages in life). Nonetheless it is both the underlying perceptions of merits and values of such transfers, as well as the ultimate effects of how they are administered that is at the nexus of political discourse. This is exacerbated in times of instability and growing perceptions of inequality within society – phenomena of current times.

Social policy thus involves multiple sectors of government and in particular: health, education, housing, migration and refugee policy, social support and advisory services, employment policy, disability and other benefits (and indirectly the tax system), justice, police, corrections, and pensions. Social expenditure typically amounts to between 30% and 50% of a liberal democracy's governmental expenditure. The provision of social services is generally a mix of directly provided services (e.g. pensions and benefits, policing and justice) and services provided by third parties (that may be local government agencies or NGOs – e.g. health, education, housing, etc.).

---

[2] Gluckman 2016: http://www.pmcsa.org.nz/wp-content/uploads/DES-Commentary-.pdf

Whereas some services are generally universal (e.g. education and health), many services are targeted. In practice the targeted services tend to engage with subgroups of the population – with a strong overlap of need and demand and complex interactions between causal factors. For example, social housing, unemployment, mental health, drug and addiction services, social welfare and justice services all show a strong overlap of provision, even allowing for common confounders like socio-economic status. Establishing entitlement criteria in these circumstances can become contentious – for example over which criteria are necessary and sufficient for the State to provide social housing.

Thus, the fundamental challenge for any government (beyond the political challenge of short-term versus long-term planning and electoral considerations) is in deciding what mix of services to provide and to whom, and at what point in their life course, include determining:

- How to balance between universal and targeted (and on what basis) programmes
- What is the marginal benefit of any particular service and how reliable is the estimate of effect size?
- Which services provided are relatively ineffective?
- How are services perceived from a client perspective?
- How to balance between preventative solutions and dealing with acute capacity and demand issues?[3]
- How to choose among competing priorities?
- How to juggle constituent demands, partisan ideology and goals, and real world fiscal and societal constraints.

The fundamental challenge for the policy maker effectively becomes: *how can investment at one point in an individual's life course create a positive rather than a negative return to the individual[4] and to the state?* In other words, how can this investment bolster the individual resilience so that they are able resist stress, remain healthy and cope with the changing demands of the labour market, rather than becoming dependent on the state? Early intervention of health, welfare and social services has the goal of reducing the risk of negative life outcomes across the lifespan. This is a long-standing well-understood and uncontested argument but the ability to translate it effectively into practice has been limited.

The general acceptability of this statement can be illustrated by a number of simple examples of intent (whether achieved or not):

---

[3] e.g. preventative health versus medical services, between educational services that might reduce crime in 20 years' time versus the need to ensure adequate policing now etc.
[4] The individual in this question can be replaced by a family or social group (e.g. a refugee group etc.).

- Universal education has as a major goal of ensuring the optimal potential of an individual and their long-term contribution as a productive member of society
- Early childhood education focused on non-cognitive skills has the goal of improving social behaviours and employability and social stability later in life
- Public health investment has the goal of sustaining the quality of life of individuals and healthy aging
- Mental health services have the goal of enabling individuals to be fully functioning and satisfied members of society
- Migration policy reflects economic and humanitarian goals but must have the goal of constructively integrating newcomers into the new societal milieu.
- Social housing policy beyond its humanitarian function has the goal of reducing other burdens on the social system (health, employment, crime etc.)
- Youth policing and justice have the goal of reducing the long-term likelihood of recidivism

The policy challenge is deciding which programmes are most likely to achieve this mix of humanitarian and social goals. The range of possible interventions is infinite and the advocates for any intervention will use a mix of normative, political, humanitarian and rhetorical arguments with very mixed access to an evidence base. This paper suggests that the use of citizen-based data alongside an appropriate scientifically coherent framework will assist policy makers in choosing among options.

A further challenge for governments in social sector policy making is the almost inevitably siloed and sometimes narrow nature of the policy process that may look neither at closely related strategies nor long-term goals. Ministries of health, for example, will generally focus first on meeting their demand side capacity needs and only secondarily on reducing that demand via public health or health promotion services solutions It is less likely that a ministry of health will advocate for high-quality early childhood education even though evidence suggests that it would, for example, reduce demand on mental health services later. Similarly the justice sector is more likely to focus on policing and prisons rather than promoting the development of executive and self-control in childhood, which has been shown to reduce criminal activity years later. Appropriate use of citizen-based analytics can allow governments to see across these silos because it can link inputs and outputs over multiple domains of interest over long time bases. Thus, if the processes of policy development extend to include **citizen-based analytics**, and provides the right intellectual framework, with databases and the associated analytic tools of sufficient quality, then cross-agency policy setting becomes more likely as a common language and a common set of metrics can emerge.

The policy framework which has been termed '**social investment**' in New Zealand 's policy framing is an attempt to address some of these challenges by combining the growing understanding of life-course phenomena with highly curated citizen-based data and analytics. In essence this provides a significant step forward in developing citizen-centric policies. It can also assist with enhancing the quality of individual–focused services albeit that there are

inherently different approaches to how science and evidence are used for policy development versus optimising service delivery.

The challenges of using scientific evidence to assist the policy process

Traditional social sciences can help address some of these questions some of the time. However the challenges of both *ex ante* or *ex post* analysis are often considered too difficult given the complexity of multiple drivers and the interactions of multiple interventions. And, when data have been used to date, they have largely been aggregate data that do not reflect the complexity of individual life journeys and key factors in those stories have generally not been available for analysis. Further, there is often a reluctance to enter into formal randomised 'policy trials' to evaluate a potential intervention, whether because of short political cycles, fears of misrepresenting particular subgroups, or concerns that randomised control trials may not be well equipped to address complex life-course problems. And, when pilot studies are conducted, there can be limitations on their scalability for many reasons. There is considerable potential benefit in having large scale, real-time data to evaluate policy impact, and/or to identify targets, which goes beyond what other research approaches can offer.

The development of appropriate big data tools can provide a conceptual way to assist with this dilemma. In theory if one knew about the multiple inputs and outcomes for individuals within a population, one could use a number of analytical techniques to see which inputs might explain significant effects on outcomes. Similarly it might be theoretically possible to identify subpopulations that respond differently. It needs to be emphasized that for a variety of reasons that this is an unrealistic and idealistic expectation because no database will ever contain information on all the biological, developmental, environmental, contextual and stochastic (i.e., random) components that affect an individual life-course. Nevertheless this hypothetical construct allows analysis of the potential value and limitations on the use of large data sets to explore social policy choices and their implementation.

Acknowledging from the outset the fundamental limitations in reaching this ideal, citizen-based data can assist significantly in moving beyond current policy tools provided the underlying assumptions are understood. Importantly the approach can guide the policy maker and analyst to those policy areas where the largest potential gains from investment for social returns or for remediation might be made.

In doing so those working with data need to have a conceptual and scientifically defensible understanding of how those inputs are likely to directly, indirectly or interactively affect a particular outcome. The approach also assumes that the inputs will have a significant effect that can be detected against the inherent noise that exists in any such data set. However it needs to be emphasized that the variation in the characteristics of people can never be gathered up in simple administrative data sets, which of necessity summarise away much

individual specificity. Administrative data sets are in general not designed to test hypotheses or model causality. Underreporting and the limited visibility of causative factors makes it important to develop systems that integrate other information sources with administrative data. Thus, not only the data management and analytical capability needs to advance, so does thinking about the nature of specific purpose collections such as statistical surveys and how to extract useful data from free form data sources.

The inputs that a government is primarily interested in are those over which it has some policy and fiscal responsibility (education, health, justice, social welfare services, etc.). However there are many other factors that affect social outcomes – some are easier to measure (e.g. family status, socio- economic status) than others (e.g. social-group interactions). But as the private sector has already demonstrated in its use of targeted electronic messaging, even limited data on an individual allows better targeting of services.

There is also a tendency to ignore what communities do, or the ways that families of diverse backgrounds function in themselves delivering services and supporting each other not only in health and care but in facilitating the education and employment of members. This diversity too needs to be captured where possible.

Ultimately government and citizens are interested in the realities on the ground; whether individuals have their needs met and services delivered in the most effective manner. Data can play an important role in assisting this need but it is important at the outset to acknowledge that the information needed to optimise service delivery at an individual level is largely distinct from that needed to the design and evaluate services at a programme level. Some of the recent debates over data privacy reflect confuse this distinction.

Progress in scientific understandings

At the same time as big data analytic capacities have developed, there have been major developments in biological, behavioural and social sciences that allow some of these assumptions to be addressed and place this complex decision-making into a coherent framework – namely the life course.

It is important to see each life as a trajectory that extends from before conception to death. It is now well established that experiences at any one point in the life-course will have echoes and consequences for resilience later in the life course and that these can be sustained by biological, behavioural/psychological or environmental (including the social environmental) mechanisms. The concept of plasticity – the way our minds, behaviours and emotions are moulded by events (both traumatic and beneficial) and our brains are rewired – describes these effects; plasticity for many dimensions is most sensitive in early life.

There is now compelling evidence that conditions in early life (extending from before conception through at least infancy) affect biological development in ways that have lasting effects on physical health, mental health and behaviour. These effects are mediated at least

in part through changes in the way genes are regulated (epigenetics) and can adversely affect the risks of developing obesity, diabetes, heart disease, asthma, depression etc.[5]

Nutrition, stress, and the psychosocial environment are the kinds of factors that modify these outcomes. Importantly the effects are not necessarily directly causal, rather they may work indirectly by altering the resilience to subsequent stressors later in life. It is also now well understood that the development of self-control and other aspects of executive function will affect the capacity to learn, to have prosocial capacities and be psychological resilient. This capacity is heavily influenced by the nature of the early environment and by the nature of family support and early childhood education in particular. Impaired executive function is later reflected in mood disorders, antisocial behaviour, impulsive and high risk behaviour, drug and alcohol abuse, impaired social relationships etc; all of these in turn are reflected in lower educational achievement, greater unemployment, less stable interpersonal relationships, mental disorder and greater likelihood of interactions with the justice sector[6]. Although developmental plasticity is greatest in early life, echoes of these types of effects, both biological and behavioural, persist well into the third decade of life.

It is clear that those who reach adulthood less resilient and less well equipped to cope with the normal stressors of social life and family relationships, have lower employability and will be more likely to develop physical and mental health disorders, to have a less rewarding life, and to be at greater risk of becoming chronically dependent on the social sector and to have more interactions with police and the courts.

At its core, the social investment model is about intervening at one stage in the life course to reduce personal, family, social and taxpayer costs later in life by promoting interventions that bolster individual resilience and/or promote more beneficial outcomes for the individual. These interventions[7] may be at any point in the life course but the intervention logic – and the impact – grows stronger the earlier in life effective interventions are made. However, the longer the gap between intervention and intended outcome the more difficult it is to make the argument and the harder it is to demonstrate an effect. This is further accentuated by the multiplicity of possible outcomes that might be impacted. The issue of proxy measures of outcomes, and the problem of finding tools that can weigh the relative effects of different interventions in nature and timing creates much of the policy difficulty.

The potential role of citizen-based analytics in providing empirical support for social policy

---

[5] Hanson MA & Gluckman PD. Early developmental conditioning of later health and disease: physiology or pathophysiology? Physiological Reviews 94: 1027-1076, 2014.
[6] Moffitt TE, Arseneault L, Belsky D, et al. A gradient of childhood self-control predicts health, wealth, and public safety. Proceedings of the National Academy of Sciences 108: 2693-2698, 2011.
[7] In many circumstances the intervention will not be individual based or focused, but family or community group based.

While this science points towards the general conceptual framework and the type/timing of interventions that may impact positively on a range of outcomes, the range of options remains large and largely normative in the absence of coincident development of big-data analytics that allow for a stronger empirical evidence base to be developed.

Medical, psychological and social science has many examples where normative arguments have been shown to be wrong despite the apparent strength and logic of claims made in the absence of robust scientific evidence. There have also been examples where the supposed evidence base has been structured along deeply normative beliefs embedding biases that were invisible before new combinations of data were brought to bear to help change the approach. Part of the latter issue can be attributed to the tendency to confuse correlation with causation and to misunderstand the direction of causality or the role of known or unknown confounders – methodological issues that can go unchallenged when results align with preconceived – if hidden – biases. All of these problems highlight the need for all analytics to be done in a way that is strongly linked to contemporary scientific understandings, which are themselves self-reflexive and proactively challenging the data for potential biases.

Governments are in increasing possession of administrative, census and operational data on their citizens and residents. Assuming social licence for its use is obtained, such data can now be analysed against appropriate models to address the questions discussed above both *ex ante* and *ex post*. Such analysis can shed light on how to choose what interventions for which populations for what effects.

The most useful data will be those that capture information on the same individuals over time and across multiple domains (health, education etc.), with as much contextual information as possible. To analyse this requires linked data sets because the data, even if available to a government come from multiple sources (administrative, census, surveys and other unstructured sources). Arguably the most important data will be to link the individual's particular social services access to their later outcomes. This means that the data have to be collected against unique identifiers to make linkage possible. Once collected however, as will be discussed below, the data uses under consideration generally only require anonymised data. While the data collected describe an individual's journey, policy relevant analysis looks at patterns across a population or subpopulation, not at the individual.

There are important caveats that relate to the range of data collected (especially client-level data about services accessed), to the quality of the data, and to the size of the evidence base. One of the emerging strengths in these data sets is that the problems of extrapolating from cross-sectional data (effectively aggregating subpopulations) effectively disappear as longitudinal data collections grow and which can then be studied in a much more sophisticated manner as a collection of life histories in which multiple (not necessarily predefined) factors contribute to the outcomes under study. Such approaches lend themselves to studying not only individual factors but also interactions between multiple factors acting across time. As such citizen-based data sets grow in depth, the capacity to

interrogate the data from the perspective of policy makers on behalf of society will become more refined. In turn, we should have a more comprehensive and deeper understanding of what is happening within our society. This should lead to more effective and evidence informed decision-making about social programmes. Done well this can also assist in informing key decisions that must be made at the provider level to optimize client experience and outcomes.

However there are also important risks and caveats that are discussed below regarding how the data are collected, curated and used. Provided these caveats are addressed and understood, citizen-focused analytics combined with the appropriate underlying scientific framework is a major step forward for enhanced social sector policy-making and delivery.

The human rights and societal interest dimension

The human rights based and moral arguments for the State provision of effective social services are clear in most democracies today. Further, the intervention logic for combining both a life-course approach with citizen-based analytics, as in the social investment approach, has an important social and human rights dimension. The life-course approach complies with the universal declaration of human rights, the UN Convention on the Rights of the Child and the general understanding by most democratic governments that a fair society requires that children must be given the optimal opportunity for reaching their full potential. However, given that the full potential is not recognised until well into adulthood, using an evidence base that links early inputs and later outcomes is a scientific expression of this rights-based approach.

When resources are limited they must be expended wisely. It is suggested that using an evidence-based approach, subject to the caveats above, is more likely than simple rhetorical argument to achieve fair outcomes. The transparency that is created by such approaches allows citizens to appreciate and see the complex rights issues and different dimensions of fairness.

Nonetheless there are challenges that must also be addressed that also have their basis in human rights. Social acceptance (or social license)[8] as to the range of uses of citizen-based analytics is required. This means that there must be clarity as to the potential and actual uses of data collected and adequate protections of the data in order to protect personal privacy through appropriate governance and oversight structures. This is a fast-moving area of social discourse and governance challenges, which extends well beyond government data collection and use. Thus social license must be continually sought and renewed through appropriate engagement and communication with relevant communities.

---

[8] Social license is in the process of obtaining, and continuing to obtain, societal consent for an action or programme: it is not a singular event but rather a continuous process.

In particular negotiating social license[9] is a process that must take into account the following considerations:

- Clearly understood objectives and intentions related to the collection of individual data. In particular, specific attention must be given to distinguishing between uses that can and should be anonymised versus those where identification of individuals will be possible or even necessary. It is important to clarify that data may need to be collected in a way that is linked to an individual so that the data set over time all relates to the same person but once collected the analysis does not need that identification to be retained. In essence, this is how government statistics units operate and these generally have a high level of trust, with the public understanding that such data sets are developed from identifiable data but that identification is then removed for all analyses. Such assurances and procedures are needed if sensitive client-level data are to be entered into citizen-based analytics for policy development and evaluation. Critically, policy evaluation will not reach its potential unless there is analyzable individual-level service data entered into databases but one cannot expect public comfort unless robust procedures to ensure that such data are then used only in anonymised ways. Using this approach programme evaluation and risk factors can be identified and personal identification is not needed. In the NZ context the integrated data infrastructure (IDI) under the custody of Statistics NZ is established with very tight access gates for policy and academic research purposes and progress is being made to ensure that use is distinct from other State data needs

- However the State may have legitimate reasons for needing to identify individuals from data – this may relate to ensuring coordinated service delivery, for public safety reasons, for case management etc. In these situations, robust processes are needed to ensure that any proposed data uses are legal, compliant, in the public interest, publically understood and secure. Such usage will need to be managed in ways that do not compromise a policy research database such as the IDI but at the same time ensure that appropriate mechanisms exist to optimize service delivery particularly in sensitive areas such as family violence where there is a need to share information between service providers but much harm could be done by poor practice breaking client confidentiality.

- It is important that there is a mutual understanding and clarity between government funding agencies, service providers, clients and indeed all citizens regarding the governance, rationale, uses and limits on the uses of data.

---

[9] In NZ this is led by the NZ Data Futures Partnership, which has undertaken public consultation on data matters. http://www.datafutures.co.nz

- Data governance must establish good practice and ensure policies over non-identified versus identified data and ensure ethical use, privacy and appropriate access[10]. There needs to be adequate attention to privacy and security and clarity as to when identification from databases is necessary. This will be the subject of a further paper in this series?

The economic dimension

The social investment model posits that appropriate interventions at one point early in an individual's life, can reduce the burden in terms of costs and harm to individual and to society later on. This of course has both social and economic benefits both to the State and to the individual. Personal and/or familial suffering, loss and compromised quality of life are legitimate costs that when taken into account alongside personal and societal economic costs, make the model even more compelling.

All Governments seek to have efficiency in their expenditures and many governments are finding that their capacity for increased expenditure in the social sector is constrained. Governments are constantly seeking to find the point of equilibrium between service provision and affordability. However it is important to distinguish between economic analysis that may help distinguish the options and using forward fiscal estimates as a proxy for benefit. This is simply because many social benefits cannot be easily monetised[11]. Estimates of forward fiscal liabilities can only been seen as very rough proxies for effect size in certain domains of cost and benefit. The lags and the difficulties of quantifying benefits in financial terms and the challenge of discounting over time mean that any estimates will always be estimates and must be seen as such. Some of the estimates will inherently have large uncertainties creating large error terms. Further, the complexity of any life-course is such that interventions may be serial (i.e., sequential through time) and benefits may be across multiple domains (and

---

[10] Work underway in New Zealand led by Statistics NZ in conjunction with the Data Futures Partnership, the Privacy Commissioner and the Office of the Prime Minister's Chief Science Advisor is currently reviewing policies and governance structures to ensure greater clarity around these issues.

[11] There are however a number of challenges which will merit ongoing econometric research. The political tendency is often to develop a common metric as a proxy of benefit of different interventions and the most likely proxy would be an econometric estimate. However, valuing human capital over a life course has proved a very difficult concept and becomes more difficult when trying to merge intangibles such as health with tangibles like earnings. In health, proxies such as DALYs and QALYs have been developed; but monetarizing these exposes the highly contentious issue of valuing a life. The greater the range of possible outcomes studied, the more complex this becomes and the underlying assumptions are likely to create significant issues.

individuals in families).[12] Citizen-based analyses allow a number of broader considerations to be taken into account[13].

Such fiscal proxies are therefore best only used to compare options within narrow framings where these assumptions are manageable. However, they may be one piece of information alongside values-based considerations used in weighing up policy options.

Data and Analytic Issues

Although it offers significant opportunities to better support government decision-making, the use of citizen-based data analytics also poses a number of challenges for governments. One challenge is to find how to build the 'system' of data, standards, tools, interpretative capacities and processes that make the best use of available resources and give all stakeholders confidence. Some of the key points (which will be expanded upon in future papers) include:

1. *Selection, collection and stewardship of data:* The establishment of a truly integrated, policy relevant database is not as simple as increasing the volume of data included in it. To be of use linked, integrated databases need:
   - a common and required set of technical standards;
   - a common set of definitions of metrics;
   - a standard format for packaging the data;

---

[12] For example, the New Zealand Government recently assessed the potential long-term fiscal savings of reducing the vulnerability of those children and young people at highest risk of long-term unemployment and poor criminal justice outcomes. After ranking all individuals from highest to lowest risk, they split the population into 20 'ventiles'. They then considered a relatively modest scenario where the risks facing each ventile were reduced to the level of risk of the next 'less-worse off ventile'. In other words, the risks for those identified as at the highest risk (100th to 95th percentiles) were reduced to the next level of risk (the 94th to 90th percentiles), and so on. That study found that under this scenario overall welfare costs would be reduced by around 12% after 40 years, and costs associated with the corrections/prison system by 21%. Similarly, they assessed a scenario where the social outcomes experienced by people living in more rural regions were improved to the level of those living in the three main metropolitan centres. That scenario was found to reduce overall welfare costs by 13%, and corrections costs by 18%. These savings would be large enough to have a tangible impact on overall government expenditure.

[13] For example the provision of social housing in New Zealand to ex-prisoners was found to reduce the spend by the Department of Corrections by 25% for those who received the support, by reducing the average length of time recipients were in prison. However, it was found to result in a 6% increase in spending on education for recipients, and a 3.6% increase in welfare payments, due to their children staying in school for longer and families accessing better support. So, in this instance, increased spending on social housing was found in net terms to lead to an increase in direct government spending in other areas, rather than a decrease. These downstream changes in education and welfare outcomes are of course positive ones. This example shows that modelling and analysis can indicate a fuller understanding of costs and benefits.

- a required and uniform process across every agency or data-collection point for data cleaning.

  This suggests that one highly trusted agency within government should be given clear responsibility for the core architecture of any integrated data system, and associated data standards. Good data management is central to the enterprise. Data management needs expertise and experience and extremely sound policy and procedure to guide it. This is a sophisticated, continuous and resource-intensive process.

There is a parallel need to give priority to developing data-hygiene, data-security and data-use skills within ministries, particularly regarding how to (and how not to) draw conclusions logically from the data that are available.

There is often limited capacity to undertake big-data analytics in the public sector and the choices of algorithms and statistics require sophisticated knowledge. A core cadre of expertise on these matters is required within government.

2. *System architecture:* There is a distinction between population level models that can inform general programme design and *ex ante* and *ex post* evaluation from that needed to manage a particular programme or an individual's entitlements, needs and services. As will be discussed below population level models and databases such as New Zealand's IDI that are needed to inform social policy development are used in a strictly anonymised manner - although key to building the database is the need to individually link data. Some programme management and client-based management needs require appropriately curated data that may or may not need to be identifiable. The issue then emerges as to whether identification is needed and where such data is held and the rules of access; this is discussed below. Despite these distinctions, the system architecture should be developed in a way to minimize duplication of effort.

3. *Data hygiene:* It is a common misunderstanding that the size of a database or sample can correct for any errors within it. This is almost never the case. For one thing, the larger the dataset the greater the likelihood of embedded errors, to which a skilled analyst must be alert. It is important to appreciate that bigger data sets do not eradicate or obviate this problem – indeed they can confound the analysis further if the data is not robust and clean. With 'noisy' data, which is often a function of poor quality control systems, important patterns will be missed because they cannot be detected. Conversely spurious associations can become reified. It is indeed possible that bad data will produce results that are the complete inverse of the actual reality.

4. *Importation of third party sourced data:* It is increasingly common in many countries, including New Zealand, for social services to be provided by third party agencies, whether NGOs or devolved branches of government. To be truly useful for policy development and programme evaluation, any integrated dataset should ideally also include point-of-care data from all service providers. However, the development of point-of-care data entry

into such databases would have a range of budgetary, policy and database hygiene implications. It may also affect social license considerations. Over time, other non-government sources of data including from the private sector would make the analyses richer but the range of social license issues would grow commensurately. In particular where services require very high trust between provider and citizen over a long period, the required protections will need to be able to manage transfer and exchanges between government and NGOs at scale.

5. *Potential data misinterpretation:* There is a range of traps that analysts of big and complex data sets can fall into, leading to potentially inaccurate conclusions. Common mistakes include: failure to understand the issues of multiple comparisons leading to false-positive conclusions; the confusion of correlation with causation; predetermination of erroneous pathways of causality; failure to include or recognise or control for confounders; and the limitations on many of the actual metrics collected, which need to be understood at the domain level. Another reason for misinterpretation are poorly formulated questions in the first instance. These issues again highlight the importance of ensuring adequate capacity and capability in public sector agencies, and potential training opportunities within universities.

6. *Algorithmic bias:* There needs to be accountability of analysis, such that potential biases built-into algorithms can be identified and eliminated or mitigated. This is an area that requires multidisciplinary engagement.

7. *Individual Prediction:* A further set of issues exists when data are used to predict outcomes at an individual level. While prediction based on risk factors is a key objective of citizen-based analytics to assist targeting interventions, it needs to be emphasized that in general such predictive approaches will identify risk and resilience factors based on group characteristics and there are significant limits and dangers in extrapolating this to a specific individual (s).

8. *Access to the data:* Clear, transparent and enforceable rules need to exist regarding access to government data – whether anonymised or not. Obviously strict processes to guard against misuse and to ensure cybersecurity must be in place. However, processes must also be in place to determine levels of access to data. The potential users extend beyond government employees. For example, academic use of such data must be encouraged but processes, such as have been developed by Statistics NZ[14], must be in place to ensure appropriate use in accord with social license, privacy rules, etc. Equally, there are valid

---

[14] In New Zealand the curation of the IDI has been assigned to the national statistics agency, Statistics NZ, which is both highly trusted and has statutory powers that ensure its independence and has highly developed rules relating to access. http://www.stats.govt.nz/browse_for_stats/snapshots-of-nz/integrated-data-infrastructure/keep-data-safe.aspx#safes

uses of public data by the private sector but again controls on access must ensure the public interest is protected.

A persistent fear is misuse of sensitive personal data. For this, appropriate legal protections are needed against such misuse in order to maintain an environment of confidence and social license. This may need to extend to consideration of making it a criminal offense to re-identify a person from an anonymised database except under specified conditions (e.g. national security), and individuals may wish to know the accuracy of their data is assured. This is not a primary issue for citizen–based analytics for policy development but arises when data are being used for individual service-delivery purposes and is beyond the scope of this paper. It is primarily a matter for those responsible for data collection and service delivery but processes may need to be developed to allow assurance on such matters.

Integration of citizen-based analytics into policymaking

There is an inherently compelling logic in combining citizen-based data with science-informed perspectives to better frame public policy approaches. The improved data and analytical approaches that have made the citizen-based analytic approach possible, allow governments to more accurately identify which types of individuals are most likely to suffer poor social outcomes over the longer term.

It also allows *ex ante* and *ex poste* evaluation of proposed or delivered programmes. Importantly it allows the complex realities of multi-dimensional inputs and outcomes over the broad span of the life-course to be properly considered. It creates the possibility of fast feedback loops to adapt programmes as information as to their effectiveness emerges. It will give greater insights into the potential of alternate approaches and choices and their relative merits as government institutions have limited bandwidth in terms of evaluating and developing potential policies and implementing new programmes. It enables governments to make a more integrated approach to policy making and, if developed well, it is likely to give citizens a greater sense that their interests are being considered holistically regarding both service provision and the use of taxpayer resources.

It needs to be emphasised that this general data-based approach is not a panacea for policy-makers, but it does allow governments to better evaluate and integrate both pre-emptive and preventative services on one hand and to enable meeting of some current needs on the other. However, not all service needs (e.g. crisis support, client management) can be addressed in this manner. Although citizen-based analytics can become an important part of social sector policy-making, they do not displace the need for continual improvement in traditional policy analysis approaches, including rigorous mixed methods and qualitative work that can expose important themes that can be formulated into questions with which to probe the quantitative data. Nor do they displace the need for continued empirical research including approaches

aimed at improving service delivery at the individual level (e.g. behavioural insights and 'what works units') and other forms of evidence-informed policy making.

To be most effective, governments will need to invest in a range of supporting activities: obtaining social license, ensuring data systems, ensuring analytical capacities and ensure that the policy process understands how science and data can inform the policy process. The better use of data and knowledge must lead to more effective services over time. It would be wrong to claim that there will be instant improvements – in most cases there is much to do to create the data systems and competencies to fully use the data in a confident and competent manner.

Such changes are also likely to have a significant impact on stakeholders. For example, in some cases the data could challenge some 'sacred cows' of public policy – that is programmes that have had strong stakeholder or political support and have become entrenched within the policy landscape but turn out to be ineffective or of low impact. When this occurs, some might seek to diminish the data and the analysis rather than accept that the programme in question is not effective (at least to a magnitude sufficient to justify continuation *vis a vis* other approaches). For this reason, it is all the more important that analysts maintain their own scepticism and question their assumptions and biases to ensure that no methodological error is at fault. Indeed, the policy community needs to invest adequately in the critical and technical skill sets necessary to support claims from the data.

At the same time political leaders will need to accept that data will potentially narrow or alter their range of options (or in some case increase them) from those that have been the basis of their traditional approaches and /or rhetoric. Because of this possibility, it is important that there are transparent public conversations about the potential of, and the limits of citizen-based data approaches. Indeed this is one purpose of this paper.

It will be important that there is continued and transparent realism about what is possible through the use of data. But gradually more refined and innovative analysis will become possible, subject to the need to be clear about data management, data curation and social license. Over time adaptive learning and new techniques will improve the ability of this approach to deliver. It will also show where the realities of individual behaviour are such that some types of outcome are not amenable to prediction or to a data-focused approach. It will show where alternate approaches to evidence gathering via policy trials are necessary.

In summary, many of the above issues will create challenges that will not always be easy for the policy process to address. It will require redirection of government resources, improved agency collaboration, changes to service delivery models, improved individual-level data, and better monitoring of service outcomes. All of these requirements can present a policy challenge. These aspects will be considered further in a subsequent paper. Some of the particular policy opportunities and challenges are outlined below.

Cross agency collaboration: As noted above, the tendency towards a siloed approach to policy making is an abiding challenge in social-sector policy development creating a dynamic that can prevent the holistic consideration of an individual. In contrast rather than seeing education, health, justice, employment etc. as separate domains with separate drivers, the citizen-based analytic approach allows policy makers to break open those silos and look at some policies in an integrated manner. It enables both short and long-term effects of any factor or intervention to be considered, with the primary outcomes often removed in time and nature from the intervention.

Further, the long time-frame of benefit can otherwise make it difficult to judge effect size both in social and fiscal terms. For example, the work of Heckman and colleagues[15] has shown that high-quality intensive early childhood education linked to intensive family support has long-term benefits on high-school completion, employment, stability of relationships and, in particular, on reducing interactions with the justice and prison systems. The highest impacts in terms of economic return on highly targeted investments are on the last of these but are not demonstrable until some 20-30 years after the intervention is made in the education and family-support domains.

Similarly, recent empirical work using the integrated data system (IDI) in New Zealand found that children exhibiting one or more of just four key characteristics were significantly more likely to experience poorer outcomes as young adults. Critically, that work also found that the number and combination of the four indicators was important. In general, the greater the number of indicators the child has the higher the likelihood that the child will have poorer outcomes. This highlights the importance of a 'whole of system' approach when developing approaches to address at risk youth. Policies relating to education, the justice sector, welfare design and child services all need to work together in concert if they are to be effective at improving the lives of the most vulnerable. This is often a challenge for agencies and may be quite a new way of working for agencies that has historically worked largely independently from each other. It in turn this will require processes by which such analyses feed into the budget and policy decision processes[16].

Allocation of resources: Science and data alone do not determine resource allocation in a democracy – there are obviously many other dimensions to government decision-making. These are reviewed elsewhere[17]. Key to resource allocation in the social sector are matters not only of long-term value but matters of fiscal prudence, social risk, immediate need, social

---

[15] see the many references at https://heckmanequation.org

[16] The NZ budget process now includes a step termed the Social Investment Panel. This panel, which is a combination of Treasury officials, social sector science advsiors, government and NGO experts, evaluates proposals against this principle and this becomes part of the formal advice to ministers.

[17] Gluckman 2013: http://www.pmcsa.org.nz/wp-content/uploads/The-role-of-evidence-in-policy-formation-and-implementation-report.pdf

acceptability, and practical amenability. Beyond these, are the realities of public and political opinions and the electoral contract. Nevertheless It would seem intuitively obvious that where governments can identify particular groups that are at a higher risk of suffering poor outcomes, that a commensurately higher level of resources should be dedicated to working with those high-risk groups. However, in some instances doing so would require a relatively sizeable re-allocation of existing resources, or at the very least ensuring the bulk of any new resource is dedicated to those priority groups or between age cohorts. Re-allocating or prioritising limited resources in this way can in practice prove difficult for governments because of stakeholder and partisan responses. Nonetheless the transparency provided by appropriate communication of the analysis that informs decisions becomes easier when it is based on a less ideological and more robust basis. Further, the improved availability of data now makes it much easier for governments to monitor the effectiveness of different policies and programmes.

Ideally, programmes that can be shown to be less effective should either be improved, or wound down. In turn, programmes that are found to be more effective should perhaps be expanded. This responsive shifting of resources between different policy programmes can be difficult for governments to achieve, even within the same policy portfolio, as closing existing programmes often results in criticism from any vested interests.

Service delivery models: There is a range of implications of this evolving approach for the ways that governments manage the delivery of social services. Perhaps most importantly, a more holistic approach becomes more possible, addressing the overall needs of the family/whanau or individual in question.

In turn this will require better coordination between service providers to ensure that their services are 'joined up' more effectively. For example some countries are now trialling the idea of establishing a single contact point that is responsible for coordinating the provision of all services to a particular family or child[18]. In addition, new contracting models and approaches to performance monitoring are also likely to be needed: these place a much stronger focus on the practical, 'on the ground' changes that have been achieved for each family or individual, rather than just enumerating the services that have been delivered. Such approaches can represent a substantial change from the traditional service-delivery models.

Programme implementation: Ultimately all social programmes are about the delivery of effective and desirable services from the perspectives of the client, society and the State. Coherence between these views is often not possible but is clearly the somewhat utopian vision of the liberal democracies. Key to this is the importance of effective implementation by service agencies, service providers and their workforces. Change in organisational culture and practice to recognise the value of citizen-based analytic insights and outcomes will take time and effort from both policy agencies and those directly involved. It will challenge some long-

---

[18] The new Ministry of Vulnerable Children-Tamariki Oranga is an example of such an approach.

established ways of working across the whole social system. It will require transparent communication in many aspects particularly related to both *ex ante* and *ex post* assessment regarding programmes and to the effectiveness of individual programmes. It offers the opportunity for adaptive learning from data. This is an area where government agencies must work with providers to ensure mutual understandings. As discussed elsewhere in this paper related issues that must be carefully and sensitively addressed include issues of non-anonymised data.

Limitations of citizen-based analytics and the social-investment approach

It is essential to understand that citizen-based analytics, while based on individual case histories, are still capable only of looking at group effects. Any extrapolation to an individual is unwise and is driven by flawed logic: the political temptation to claim precision of individual prediction should be avoided.

For example, while it is true that such analytics can help identify particularly at-risk individuals with greater precision, those outside of that group still have a non-zero risk of suffering poor outcomes. Moreover, because the latter come from a much larger population overall, the number of individuals from that lower risk group who go on to experience poor outcomes can be quite large. This is illustrated by a New Zealand study using the IDI database[19] which found that while 20.4% of those with two or more of the four risk indicators are expected to spend more than two years on social benefit before age 21, compared to 4.4% of those with none or one indicator, the absolute numbers of children from each group spending time on a benefit before the age of 21 are broadly similar; 10,800 from the high-risk group and 13,700 from the lower-risk group. This highlights the fact that the data-driven approach is not a 'silver bullet'. Policies targeting high-risk youth should be seen as a valuable complement to broader whole-of-population policies, not a replacement for them.

More generally, although analyses of the data to date has provided substantial insights there is still much that we do not fully understand. Further, steady improvements in our understanding are likely, through for example, the research partnership between the New Zealand Government and the European Commission's Joint Research Centre, and as more data

---

[19] Recent empirical work in New Zealand found that children exhibiting one or more of the following four key characteristics were much more likely to experience poorer outcomes as young adults: having a finding of abuse or neglect, or having spent time in care of child protection services; having spent most of their lifetime in a household supported by benefits; having a parent who has received a community or custodial sentence; having a mother who has no formal qualifications. Children exhibiting two of more of these indicator characteristics were found to be three times more likely to leave school with no qualifications, six times more likely to receive benefits for more than two years before the age of 21, and ten times more likely to spend time in jail before the age of 21. Perhaps more importantly, the evidence also suggested that to at least some degree well designed and targeted interventions in the early years of a child's life can help to shift the entire trajectory of their adulthood.

become available, and more analysis is undertaken. New analytical tools and skills will need to be developed, yet a degree of uncertainty seems certain to remain. The design and implementation of social policies will therefore always require use of expert judgement. Moreover, policy mistakes will almost certainly be made, data driven or not.

Conclusions

At its heart the citizen-based analytics to support a 'social-investment' approach to policy-making has a simple but powerful goal; to develop each generation of citizens to its fullest possible potential, bolster their resilience and subsequently to sustain their quality of life. As the State has a responsibility to fund programmes that achieve this, it is in the interests of all citizens that the most effective programmes are funded. While there will always be political and ideological judgement as to what are the priorities and particular goals and range of acceptable approaches[20], science-informed data analytics provide a major step forward in ensuring that these judgements are informed by the best evidence as to what works for whom and under what circumstances, what does not work, who may be harmed, what are the spill-over benefits and costs and what is the relative effect size of any proposed approach.

Additionally such a focus will have spill-over benefits for other forms of evidence gathering and policy development and evaluation (such as continuous improvement and adaptive learning approaches).

Although the citizen-based analytics for social-investment approach is about improving the lives of all citizens, it has a particular focus on the most vulnerable, and on ensuring greater resilience by taking a life-course approach. Disadvantaged citizens who have very complex problems often only come to the attention of social-service agencies when something really goes wrong. The social-investment approach looks to identify and address these problems at an earlier, more tractable stage, and more effectively, and to develop a long-term strategy, as opposed to focusing only on a reactive response.

The reality of governments however is that, notwithstanding the promise of such approaches, there is and will always be a need (which hopefully will reduce over time) to provide services for those in need of immediate assistance. Appropriately framed data can help identify the optimal and the most effective programmes for those individuals and allow for integration at an individual level based on individual need.

At a conceptual level none of this is particularly new. What has changed is the ability of governments to utilise a more data driven approach. Until now most of the arguments that underpin social policies have been largely normative and rhetorical.   Improved

---

[20] For example if the data suggest that lengthy remand in prison, rather than on bail, for a youth offender is more likely to lead to recidivism and a life of crime. This is borne out by a sufficient number of case examples to confront some well entrenched popular and political positions.

understandings in developmental, behavioural and social science combined with rapidly evolving big-data methodologies mean we can now better identify who needs more help and when. This allows us to understand much more clearly how specific government (and non-government) services impact on all of the outcomes that matter to a person over their lifetime. The core criteria for introduction of this approach include:

- Scientific advice that ensures the models developed are biologically, behaviourally and socially plausible and operationally realistic.
- Social license to use citizen data for policy research and evaluation (note there is a real distinction here from the use of citizen data by governments in other ways: e.g. regulatory compliance issues).
- A transparent and trusted process for curating citizen data from multiple sources including longitudinal data on individuals.
- Access to modelling tools for analysing big data
- Individuals trained in data curation, management and analysis.
- A policy and political community that understands both the opportunities and limitations of citizen-based analytics.

As this paper points out, considerable culture change is needed for this type of approach to be successful and meet its promise. Firstly there needs to be broad transparency and understanding regarding the intervention logic and the consequential use of scientific underpinnings to inform data analytics in social policy making. Secondly there needs to be attention paid to the governance, management and stewardship of data and clarity and accountability as to purpose of collection and use. There is value in creating as full a data picture as possible, which includes data from both within and beyond governments, but there must be clear boundaries as to uses that rely on identifiable data versus anonymised data. Thirdly there needs to be an understanding of the relationship between citizen-based analytics and other aspects of social policy development and implementation. Fourthly culture change and clarity of understandings is needed at every level: from clients, service providers, policy makers and politicians. Fifthly these new approaches will require governments to be willing to accept findings which may be inconvenient in political or ideological terms because they will suggest the value of shifting resources around the system as social insights are developed on a robust basis. Only then can these important developments in understanding the lives of citizens translate into more effective policies that could benefit the whole of society.

There will be ongoing learning as governments such as New Zealand's develop citizen-based analytics for social policy. The intent of this paper has been to summarise how such approaches can be understood at this stage in their development. Hopefully it can assist conversations on what are important but complex matters. As experience is gained with such approaches, no doubt they will evolve and adapt.

ACKNOWLEDGEMENTS