Tim Hughes

Predictive Analytics for Policy and Practice

reflections from the criminal justice system

The current government is seeking to take the logic of an investment approach to welfare and apply it to other areas of expenditure. Like most sectors, the justice sector has a programme of work underway to improve its ability to make good investment decisions; in a justice sector context this primarily means applying resources where they can best reduce the long-term social and economic costs of crime.

Good investment decisions are more likely made when decision-makers are provided with quality data analytics. In particular, there is growing interest in the use of administrative data to predict which individuals, families and communities are at high risk of ongoing dysfunction, to help the targeting of resources to areas of greatest need. In this article I use examples from the criminal justice system to offer some practical suggestions for how the promise of better predictive analytics can be pursued across government while also minimising the risks. I make six suggestions:

- be specific when communicating predictive statements to decision-makers;
- consider prediction and response together;
- carefully consider the optimal degree of targeting for any given service;
- minimise any unintended side effects of acting on the basis of predictions;
- consider predictive analytics in the context of a wider information strategy;
- consider prediction in the context of a wider practice framework.

While the examples are drawn from criminal justice, the lessons are quite general and may help with the broader move towards greater use of predictive analytics in government. For example, they may be useful to analysts who are looking to exploit the opportunities provided by the Integrated Data Infrastructure at Statistics New Zealand, or those identified by the New Zealand Data Futures Forum.

There is little doubt better use of data with predictive analytics will support the development of better evidence-informed policy and practice, but there are also several risks with this approach. Some of these are familiar, such as risks to privacy.

Tim Hughes studied economics at the University of Auckland and has worked in policy roles at the Department of Corrections and the Ministry of Social Development. He is currently working in the sector investment team at the Ministry of Justice. The views in this article are those of the author and do not necessarily reflect those of the Ministry of Justice.

Table 1

	No conduct problems at age 7–9	Severe conduct problems at age 7–9
Not arrested or convicted between ages 18 and 30	453 (89%)	47 (48%)
Arrested or convicted between ages 18 and 30	55 (11%)	49 (52%)
Total	508 (100%)	96 (100%)

Others are less often discussed, such as the simultaneity between prediction and response. In this article I elaborate some of these less familiar risks and suggest how to mitigate them.

Suggestion one: be specific when communicating predictive statements to decision-makers

Some predictions are more accurate than others. This is intuitively obvious when we consider the accuracy of the weather forecast one, five or ten days into the future, and the difference between predicting high winds in general over a large area over an afternoon versus predicting a specific wind speed at a specific location at a specific time. Predictions relevant to policy and practice can differ in their accuracy just as much as predictions about the weather, but this subtlety can sometimes be unintentionally lost in communication to policy makers and practitioners.

For example, consider the familiar observation that adult offending has its roots in childhood disadvantage. This link has been explored using data from the Christchurch Health and Disability Study, a study of 1,265 people born in Christchurch in 1977. Using data which tracks these individuals to the age of 30, Jakobsen, Fergusson and Hallwood (2012) report that there are 'significant associations between early conduct problems and later crime'. Table 1 above summarises the data supporting this result.

It is clear from this table that a much higher proportion of children with severe conduct problems (52%) went on to offend as adults compared to children with no conduct problems (11%). This is the 'significant association' reported by the researcher. However, this link is far from determinative. It is striking that among those who were arrested or convicted between the ages of 18 and 30, more had no conduct problems as children (55) than had severe conduct problems (49). This implies that targeting children with severe conduct problems has the potential to address at most half of crime in adulthood.

Further, nearly half of children with severe conduct problems did not offend at all as adults. And note that 'any arrest or conviction' is a very low bar for describing someone as an offender; many of the arrests would likely have been one-offs, for petty offences. These subtleties may be lost if non-technical readers look at the phrase 'significant association between conduct problems and later crime' but are not also provided with information about the strength of that association. This is not a criticism of the Jakobsen et al. research; merely an example to illustrate how communication between researchers and decision-makers can be challenging in the area of risk prediction.

A key challenge for communication is to express predictions as fundamentally probabilistic and uncertain. In the example of the children exposed to risk factors during childhood, whether or not those risk factors go on to be manifested as offending later in life depends on a host of contingencies, such as the child's degree of success in education, the peers they associate with during adolescence, whether they develop a substance abuse problem, whether they succeed in finding a job, what community they end up living in, and so forth (Farrington and Welsh, 2005; Pratt and Cullen, 2000).

This is a general problem in criminal justice. Prediction is always imperfect, because the causes of crime are too complex to be fully specified in a statistical model. Crime does not flow deterministically from criminal traits, but is rather the product of a complex and stochastic interplay of personal, situational and other contextual factors, such as peer relationships. A recent systematic review of empirical tests of theory in crime and justice found that few studies managed to explain more than a small proportion of the variance in crime (Weisburd and Piquero, 2008). And, unfortunately, it is often the case that the outcomes we would most like to predict are those that are hardest to predict accurately. For example, rare crimes such as murder and rape cause more harm than common offences such as car theft, but the very rarity of these most serious crimes makes them much harder to predict. Predictability depends crucially on base rates, with low base-rate events being intrinsically difficult to predict (Gottfredson and Moriarty, 2006).

In a low base-rate context, even a statistical tool that improves substantially on the base rate can still have limited practical value. For example, a new tool was recently developed to support parole officers in Pennsylvania (Berk et al., 2009). The tool predicts the risk of murder for parolees. 1% of parolees in Pennsylvania are charged with murder within two years of release. The tool is able to improve upon this base rate of 1%, but among those classified 'high risk', only 7% were actually charged with murder – a false positive rate of 93%. And more murder charges (185) were laid against people in the 'low risk' group than in the 'high risk' group (137). In this case, communicating that someone is of 'high risk' of committing murder may be misleading, if it leads someone to inaccurately believe that that individual is 30%, 50% or even more likely to commit murder.

While prediction is particularly difficult in low base-rate situations, it is also true that general offending cannot be predicted beyond a certain level of accuracy. As summarised in a recent meta-analysis of risk analysis tools in criminal justice, 'the view that violence, sexual or criminal risk can be predicted in most cases is not evidence based … risk assessment tools in their current form can only be used to roughly classify individuals at the group level, and not to safely determine criminal prognosis in an individual case' (Fazell et al., 2012, p.5). It is this intermediate level of knowledge that can pose the most difficulty for decision-makers. Often we are in the uncomfortable position of knowing something more than nothing, but much less than everything; knowing precisely where on the spectrum of uncertainty we lie in any given predictive context is essential to making good evidence-based decisions.

Suggestion two: consider prediction and response together

One of the key intuitive promises with predictive analytics is that with a better understanding of where adverse outcomes are concentrated, government will better be able to target support where it is most needed. But it is not always the case that those we identify as being most likely to experience adverse circumstances are those on whom effort is best concentrated. While predictive analytics are useful for this kind of decision, they are arguably less important than assessments of what Harcourt (2007), adopting the language of economics, calls elasticity, and what correctional psychologists call responsivity: that is, sensitivity to intervention.

For example, psychopathy is an important risk factor for reoffending, but there is evidence to suggest that rehabilitation programmes are often less effective for psychopaths (Tew, Harkins and Dixon, 2013). If we focus only on the risk, and not on the amenability to change of someone with a high psychopathy rating, we might offer the individual a rehabilitation programme to no effect, whereas a somewhat lower-risk individual might have gained from treatment. Public safety, in other words, might have been better served by targeting support to the lower-risk individual.

Ultimately, the effectiveness of intervention with a given group is a separate empirical question that predictive analytics is often not well suited to answer. To continue the example of childhood conduct problems from above, even though there is a correlation between childhood conduct problems and later offending, it does not immediately follow (even though it is intuitively reasonable) that by addressing conduct problems, later offending will also be reduced. In fact, the limited evidence we have to date on this topic suggests that while early intervention can reduce later offending on average, it does so by a relatively modest amount. The meta-analysis undertaken by Dekovic et al. (2011) found an average effect size of intervention before age 12 on offending after age 18 of 1.26, expressed as an odds ratio. This means that early intervention with a group of children among whom 20% would have offended as adults can reduce the proportion who actually offend to 17%. This is a real reduction, but is smaller than that which can be achieved by interventions later in the life course, such as alcohol and drug treatment (Mitchell, Wilson and Mackenzie, 2012).

The most useful risk prediction is unlikely to be about merely sorting the good from the bad apples, but rather about understanding complex patterns of continuity and change, turning points, and appropriate interventions.

Considering prediction and response together also helps us to reconceptualise risk prediction in a more productive way. The most useful risk prediction is unlikely to be about merely sorting the good from the bad apples, but rather about understanding complex patterns of continuity and change, turning points, and appropriate interventions. For example, we would value a tool that specified the best time for a given type of intervention, given previous history of intervention, recent changes in family situation, etc, or to help guide decisions about sequencing of interventions. These are fundamentally different predictive questions to those posed by an exercise to merely identify the most prolific and serious offenders.

Another important reason to consider prediction and response together is that social processes, including crime, do not operate in a closed system. Crime leads to responses by government agencies that can themselves affect crime, for both good and bad. Where these responses are taken on the basis of actuarial predictions, these reactions can in turn change the pattern of crime, thus undermining the accuracy of the prediction tool that led to the response.

There are two different problems these reactions can cause: one where interventions reduce risk; and the other where interventions intentionally or inadvertently increase risk. In most cases, interventions are designed to reduce risk. For example, the risk profile of the prisoners released from prison each year is lower than it would be if it were not for the delivery of services such as alcohol and drug treatment, which have good evidence to show that they reduce the likelihood of reoffending (Department of Corrections, 2013; Mitchell, Wilson and Mackenzie, 2012). In these cases, the problem is that it becomes more difficult to differentiate between prisoners on the basis of observed reoffending.

So, in some cases, poor predictive accuracy may be evidence of an effective system. For example, the perfect parole system would be one that had a fixed threshold of 'acceptable risk', and released everyone as soon as they reached that threshold and not before. In this situation, everyone released on parole would pose exactly the same amount of risk, so would be indistinguishable by statistical analysis. In this sense, Bushway and Smith (2007) argue that an inability to discriminate between risk levels can be a measure of successful performance, as it indicates that all available information is being used efficiently.

The perhaps more troubling case is where the response to risk causes that risk to increase. There are several examples of this in the justice system. For example, there is some evidence that placing prisoners in high-security facilities can increase their likelihood of committing misconducts (Chen and Shapiro, 2007; Jonson, 2010). It is necessary to place certain individuals in high-security conditions to help manage the consequences of misconduct, such as by making it more difficult for prisoners to access or manufacture weapons. But when these more stringent conditions lead to an increase in the likelihood of assault, then when officials next come to calibrate the security classification system they will find that the level of risk among those in higher security has increased. In other words, the characteristics of those prisoners will thus be reinforced in the classification system. This then makes others with those characteristics more likely to be placed in high security, which then elevates their risk, and so those characteristics get an even higher loading at the next recalibration, and so on. By this stage the prediction tool may accurately reflect actual problem behaviour, but may also, to the extent that the problem behaviour is caused by placement in high-security conditions, be predicting that people with those characteristics will be placed in high security.

This circularity is described by Harcourt (2007) as a 'ratchet effect', where initial small differences in risk are inflated over time by the way the state responds to those risks. The quiet danger of the ratchet effect is that it is largely invisible. With a well-calibrated tool, those identified as high risk will indeed offend at a higher rate, thus providing clear justification for the tool. But because of the simultaneity between risk assessment and intervention, it is very difficult to separate out the 'baseline' or 'natural' risk of the individual from the 'state-generated' risk that results from responses to that risk.

More generally, the final reason to think about prediction and response together is that in a small country our analytical resource is precious. Predictive analytics on their own do not lead to better outcomes. In my view, questions of what approaches are most effective should generally take primacy. I see the main value of predictive analytics being as a tool to help consider how to take approaches that have been evaluated as effective and to make them more effective, through improving our ability to target those services to those who benefit most from them.

... police interventions focused on high-crime offenders tend to be less effective than interventions focused on high-crime places, perhaps because place-based prediction is more stable over time than offender-based prediction

Suggestion three: carefully consider the optimal degree of targeting for any given service

Predictive analytics naturally lend themselves to the targeting of support. But more specific predictions are often less accurate. For example, it is easier to predict that Japan will have a certain number of big earthquakes over the next 100 years than to predict that a 7.2 magnitude earthquake will strike tomorrow at 7am three kilometres west of Kobe. Similarly, predictions about crime tend to be less accurate the more they move from groups to individuals, from general offending to particular types of offending, and so on.

This suggests that it is not safe to assume that more highly targeted services will always be more effective, if it is not in fact possible to target intensive services with sufficient accuracy. The more highly targeted an intervention, the more likely it is to capture only a portion of those people, areas and so forth that will ultimately be involved in crime. The appropriate mix between more and less targeted interventions is ultimately an empirical question which needs to be determined on a case-by-case basis, taking account of the number of false positives and false negatives, the cost-per-target of the intervention, and the total number of people being targeted. We have already seen with the early intervention example that very early intervention can suffer from the limitation of a high number of false positives.

To take another example, we know that police interventions are more effective where they are highly targeted, but also that police interventions focused on high-crime offenders tend to be less effective than interventions focused on high-crime places, perhaps because place-based prediction is more stable over time than offender-based prediction (Lum, Koper and Telep, 2011). For example, where alcohol-related violence is a problem in an entertainment district, managing serving practices in the local bars may well be more effective than trying to directly manage the offenders whose risk becomes elevated when they drink in town, but who otherwise will often not pose a major risk. Further, evidence on school-based crime prevention suggests that the most effective approach is for coordinated behavioural management across the whole school; in the absence of this environmental change, targeted approaches for high-risk students appear to be mostly ineffective (Gottfredson, Wilson and Nejaka, 2002).

These examples also show how targeting based on predictive analytics alone can lead to a policy response that appears sensible but is less effective than an alternative – focusing on high-risk individuals instead of places, in the policing context, or instead of the school environment in the schooling context.

Similarly, in many cases there will be a trade-off between accuracy and the amount of potential harm available to be avoided, as illustrated in Table 2. It is always tempting to think that by intervening earlier, or targeting effort based on the projected outcomes over a period of many years, we will be able to address concentrated problems more effectively. But in many cases this broader view may come with a cost: a reduction in accuracy, and a reduced ability to target with precision. For example, in the case of when in the life course to intervene, the ideal balance of intervention is likely to be composed of relatively broad interventions early on, with a growing level of intensity as our ability to predict residual life-cycle offending grows. The exact balance is, of course, an empirical question, and again one that predictive analytics cannot answer alone in the absence of evaluation evidence.

Table 2

Prediction window	Long-term (e.g. many years)	Medium-term (e.g. a few years)	Short-term (e.g. a few months)
Accuracy	Lower accuracy	Medium accuracy	Higher accuracy
Precision level	Broad groups, general locations	Smaller groups, specific locations	Small groups, some individuals
Associated intervention type	Primary prevention	Secondary prevention	Tertiary prevention
Potential harm available to be avoided	Higher potential	Medium potential	Lower potential

Suggestion four: minimise any unintended side effects of response

Prediction of risk implies a duty to respond. Where that response can or does cause harm, this raises important ethical and practical implications. There is no shortage of examples in criminal justice of responses causing harm. Most directly, sentences are often intended to cause harm of a sort by restricting the liberty of those convicted of a crime. This may be justified on the grounds of just desserts, but where sentences and orders are imposed or modified on the basis of predicted risk, such as with preventive detention sentences, and with parole and bail decisions, the potential for ethical dilemmas where predictions are less than perfect is much greater.

In addition to these direct harms that are experienced by the offender, there is international evidence that certain types of justice system responses can lead to further victimisation, as they have the potential to increase the likelihood of reoffending. For example, a recent international meta-analysis suggests that police decisions to send young offenders to court can increase their subsequent likelihood of offending (Petrosino, Turpin-Petrosino and Guckenburg, 2010). Similarly, there is some international evidence that incarceration can in some cases increase the likelihood of future offending (Jonson, 2010), particularly for young offenders (Lambie and Randell, 2013), and that intensive supervision can increase reoffending compared to regular supervision (Cochran, Mears and Bales, 2014). Certain forms of intervention intended to reduce reoffending, such as Scared Straight programmes, actually increase it (Petrosino, Turpin-Petrosino and Buehler, 2013).

In addition to these obvious harms, there is a range of other areas where targeting based on predictive analytics can potentially have adverse side effects. For example, the potential for stigmatisation is greater with more targeted approaches, and may encourage essentialist thinking on the part of front-line practitioners, by which I mean thinking that focuses largely on individual propensities and neglects broader situations and contextual factors that can also drive or mitigate risk. Targeting based on risk may also have the potential to undermine therapeutic relationships between practitioners and clients, particularly if the service is mandatory. Whether any response to predicted harm is voluntary or not is also an important consideration, because in many cases a service that is effective if voluntary, such as restorative justice, becomes ineffective if mandated (Bonta et al., 2006). Targeted attention can lead to hypervigilance, with greater likelihood of offences being captured and recorded, and the related problem of the ratchet effect, discussed earlier.

The potential to cause harms is partly a practical matter, but is perhaps more importantly an ethical issue. An important ethical consideration will often be the effect of the policy or practice response on the false positives – those wrongly identified as at high risk. In a recent review for the Ministry of Social Development of ethical considerations when applying risk prediction to prevent child maltreatment, Dare (2013) noted the importance of minimising the burdens placed on those identified as high risk as a prerequisite to ethical use of predictive analytics in that context.

It is difficult to generalise about the ethical use of predictive analytics because the appropriateness depends on the context in which the prediction is applied, as well as the accuracy of the prediction. For example, consider the Berk et al. (2009) tool for predicting the likelihood of murder described earlier. This tool identified a group of parolees among whom 7% could be expected to commit a murder. This tool is used in Pennsylvania to determine the degree of parole supervision placed upon each individual. In this case, the burden placed on the 93% of people falsely identified as high risk for murder is fairly low in comparison to the value of preventing so many murders. In contrast, if this tool were used to inform parole decisions or home detention sentences, then the burden shouldered by the 93% of false positives becomes a more difficult thing to weigh against the murders prevented by incapacitating the 7% of true positives.

Suggestion five: consider predictive analytics in the context of a wider information strategy

The accuracy and appropriateness of predictive analytics depend crucially on the underlying data. If data are inaccurate, biased or incomplete, then predictive analytics may not only be limited in usefulness, but may be misleading.

Prediction can only be as good as the underlying data. This point almost goes without saying, but is worth reinforcing when it comes to the analysis of government data sets, most of which are developed for administrative purposes. Even basic pieces of data like dates of birth can be inaccurately reported, particularly to justice sector agencies, and errors in data recording can multiply when various data sets are connected together with behind-the-scenes algorithmic matching. These data errors compound on top of the general limits to prediction outlined above, adding a further source of inaccuracy.

Another important limitation is that government data sets capture variables that may be only loosely related to the outcome of interest. For example, two of the factors that are most predictive of offending are anti-social attitudes and anti-social peer groups (Andrews and Bonta, 1998). Neither of these factors is likely to be captured reliably and comprehensively in any government database. Correlations with variables such as age and number of court appearances will often be statistically significant, but in many cases these variables will be less accurate predictors than detailed psychological, socio-economic and behavioural data.

In general, we can also expect government data sets to suffer from systematic measurement error, because the data is collected for administrative, not research purposes. We only capture data on people we have to transact with, and only at the points at which we transact with them. We know, for example, that most crimes are not reported (Ministry of Justice, 2009), and that crimes committed by Māori are more likely to be reported (Department of Corrections, 2007). So any success in predicting crime as recorded in official databases may to some extent simply be success in predicting patterns of reporting, rather than in predicting crime. Similarly, there is evidence that sentencing practice for driving offences varies substantially across court districts in New Zealand (Goodall and Durrant, 2013), so risk differentials for offenders across the country will to some extent reflect differences in court practice.

All of this suggests that a strategy of increasing reliance on predictive analytics may need to be matched by considered investment in data capture and storage. The Performance Improvement Framework results show that most government agencies rate as 'needing development' when it comes to information management (Te Kawa and Guerin, 2012), suggesting that greater use of predictive analytics will need simultaneous investment in improving data management in most agencies.

> Predictive analytics will help us improve government services and make more of a difference to people's lives. At the same time, predictive analytics has limitations.

Given that quality analytics need investment in quality data, it seems important to consider the expense of improving government data sets in comparison with funding other forms of data collection. For example, it may be more cost-effective, and more informative, to fund the more detailed data collection of a quality longitudinal study, like the Christchurch Health and Disability and Growing Up in New Zealand studies. Studies such as these are custom-built to capture complex patterns of continuity and change over the life course and identify the most important static and dynamic drivers of risk. Given the inherent limitations of administrative data, longitudinal research may provide the best way to improve our ability to understand deep patterns of behavioural development and prediction.

Suggestion six: consider prediction within the context of a wider practice framework

Predictive analytics can support strategic decision-making by ministers and senior managers, but in many cases their main value may be in supporting decisions on the front line. In the criminal justice system this can include decisions such as whether to proceed with a prosecution, whether to grant bail and whether to grant parole. In many of these situations, actuarial decision-making tools may have the potential to support better decision-making. An extensive research history consistently shows that while unstructured risk assessment typically performs better than chance, actuarial risk assessment consistently performs better than either (see, for example, the meta-analyses of Aegosdottir et al., 2006 and Grove et al., 2000), even if it is usually far from perfect.

Actuarial decision-making may also be a cost-effective approach in some cases. It is unlikely that we would ever want to automate parole decision-making, but where lower-stakes decisions can be accurately made with an algorithm, such as the degree of attention paid by the Department of Corrections to various people with unpaid fines, then algorithms may be more cost-effective than professional judgement.

There are also situations where structured tools can help address concerns with bias or discrimination. For example, there is some evidence that for any given offence, and after controlling for factors such as age and offence history, Māori are somewhat more likely to be apprehended, prosecuted and so on (Department of Corrections, 2007). This may be for many reasons, but it is possible that patterns in the use of discretion by practitioners are a contributing factor. In this case, structured decision-making can make transparent the grounds for treatment, and help to address any real or perceived subjective bias in decision-making.

At the same time, the value of better prediction depends intricately on how that prediction is understood by practitioners and used to inform decision-making. There is some risk that more accurate prediction may be embraced by practitioners where it tallies with their subjective judgements, but may be disregarded otherwise. For example, Miller and Maloney (2013) found that practitioners' decision-making can be unrelated to a risk tool's result, even when the risk tool is filled out accurately. As such, predictive analytics will only generate value to the extent that the people who use it have the appropriate skill level to engage critically with the relevant findings. A structured professional judgement approach implies a high degree of skill on the part of decision-makers. If a professional is going to override the decision implied by a well-validated predictive tool, then it seems reasonable that they should hold a high level of understanding of the factors driving the predictive tool and its limitations.

This suggests that a sensible approach when using predictive analytics to inform practice decisions would be to regularly record and monitor the use of overrides. It may also be useful to provide feedback to individual practitioners about their use of overrides and the outcomes for the people they have chosen to exercise an override. Similar to the requirement for anaesthetists to follow-up with patients to identify any side effects and help them to develop a good intuition for risk, intuitive decision-making is most likely to become accurate over time with repeated exposure to the consequences of decision-making (Klein, 1998).

Conclusion

Predictive analytics will help us improve government services and make more of a difference to people's lives. At the same time, predictive analytics has limitations. Because of the technical limitations of prediction, and because the practical and ethical implications of these limitations are often so substantial, a serious and ongoing discussion about predictive analytics among policy makers, advisors and practice leaders seems warranted.

In this article I have limited discussion to criminal justice. The trade-offs in criminal justice are often very substantial, such that many of the examples in this article will not be as directly relevant to more prosaic areas of policy. But high-profile uses of prediction, such as in criminal justice, will also tend to attract greater scrutiny, and are often associated with extensive procedural protections, such as in court proceedings. Simply because the stakes are so high for both individual liberty and the safety of the community in decisions such as whether to impose a preventive detention sentence, the limitations of prediction are understood by those working in criminal justice perhaps as much as anyone. As government seeks to extend the use of predictive analytics to other parts of its business, where the procedural protections are often lower and the number of people affected by policy decisions often greater, the body of knowledge in criminal justice will perhaps offer insights for other areas where the limits of prediction are not as salient in policy discussions, and contribute to a more general discussion about predictive analytics.

Acknowledgements

Many thanks to the two reviewers for their helpful comments, and also to Paul Henderson of the Ministry of Justice, Peter Johnston of the Department of Corrections, Tony Burton and Nick Carroll of Treasury, and especially Ross McKay of the Ministry of Social Development for extensive comments on an earlier draft.

References

Aegosdottir, S., M. White, P. Spengler, L. Maugherman, R. Cook, C. Nichols, G. Lampropoulos, B. Walker, G. Cohen and J. Rush (2006) 'The meta-analysis of clinical judgement project: fifty-six years of accumulated research on clinical versus statistical prediction', *Counselling Psychologist*, 34 (3), pp.341-82

Andrews, D. and J. Bonta (1998) *The Psychology of Criminal Conduct*, Cincinatti: Anderson

Berk, R., L. Sherman, G. Barnes, E. Kurtz and L. Ahlman (2009) 'Forecasting murder within a population of probationers and parolees: a high stakes application of statistical learning', *Journal of the Royal Statistical Society*, 172 (1), pp.191-211

Bonta, J., R. Jesseman, T. Rugge and R. Cormier (2006) 'Restorative justice and recidivism: promises made, promises kept?', in D. Sullivan and L. Tifft (eds), *Handbook of Restorative Justice: a global perspective*, London: Routledge

Bushway, S. and J. Smith (2007) 'Sentencing using statistical treatment rules: what we don't know can hurt us', *Journal of Quantitative Criminology*, 23, pp.377-87

Chen, M. and J. Shapiro (2007) 'Do harsher prison conditions reduce recidivism? A discontinuity-based approach', *American Law and Economic Review*, 9 (1), pp.1-29

Cochran, J., D. Mears and W. Bales (2014) 'Assessing the effectiveness of correctional sanctions', *Journal of Quantitative Criminology*, 30, pp.317-47

Dare, T. (2013) *Predictive Risk Modelling and Child Maltreatment: an ethical review*, paper prepared for Ministry of Social Development, Auckland: University of Auckland

Dekovic, M., M. Slagt, J. Asscher, L. Boendermaker, V. Eichelsheim and P. Prinzie (2011) 'Effects of early prevention programmes on adult criminal offending: a meta-analysis', *Clinical Psychology Review*, 31 (4), pp.532-44

Department of Corrections (2007) *Over-representation of Māori in the criminal justice system*, Wellington: Department of Corrections

Department of Corrections (2013) *Annual Report*, Wellington: Department of Corrections

Farrington, D. and B. Welsh (2006) *Saving Children from a Life of Crime*, Oxford: Oxford University Press

Fazell, S., J. Singh, H. Doll and M. Grann (2012) 'Use of risk assessment instruments to predict violence and antisocial behaviour in 73 samples involving 24827 people: systematic review and meta-analysis', *British Medical Journal*, 345, July

Goodall, W. And R. Durrant (2013) 'Regional variation in sentencing: the incarceration of aggravated drink drivers in the New Zealand district courts', *Australian and New Zealand Journal of Criminology*, 46 (3), pp.422-47

Gottfredson, D., D. Wilson and S. Najaka (2002) 'School-based crime prevention', in L. Sherman, D. Farrington, B. Welsh and D. MacKenzie (eds), *Evidence-based Crime Prevention*, London: Routledge

Gottfredson, S. and L. Moriarty (2006) 'Statistical risk assessment: old problems and new applications', *Crime and Delinquency*, 52, pp.178-200

Grove, W., D. Zald, B. Lebow, B. Snitz and C. Nelson (2000) 'Clinical versus mechanical prediction: a meta-analysis', *Psychological Assessment*, 12 (1), pp.19-30

Harcourt, B. (2007) *Against Prediction: profiling, policing and punishing in an actuarial age*, Chicago: University of Chicago Press

Jakobsen, I.S., D. Fergusson and J.L. Horwood (2012) 'Early conduct problems, school achievement and later crime: findings from a 30-year longitudinal study', New Zealand Journal of Educational Studies, 47 (1), p.123-35

Jonson, C. (2010) 'The impact of imprisonment on reoffending: a meta-analysis', doctoral dissertation, University of Cincinnati

Klein, G. (1998) *Sources of Power: how people make decisions,* Cambridge, Mass: MIT Press

Lambie, I. and I. Randell (2013) 'The impact of incarceration on juvenile offenders', *Clinical Psychology Review*, 33 (3), pp.448-59

Lum, C., C. Koper and C. Telep (2011) 'The evidence-based policing matrix', *Journal of Experimental Criminology*, 7 (3), pp.3-26

Miller, J. and C. Maloney (2013) 'Practitioner compliance with risk/needs assessment tools: a theoretical and empirical assessment', *Criminal Justice and Behavior*, 40 (7), pp.761-37

Ministry of Justice (2009) *New Zealand Crime and Safety Survey 2009*, Wellington: Ministry of Justice

Mitchell, O., D. Wilson and D. Mackenzie (2012) *The Effectiveness of Incarceration-based Drug Treatment on Criminal Behaviour: a systematic review*, Campbell Systematic Reviews, 2012: 18

Petrosino, A., C. Turpin-Petrosino and J. Buehler (2013) *Scared Straight and Other Juvenile Awareness Programs for Preventing Juvenile Delinquency*, Campbell Systematic Reviews, 2013: 5

Petrosino, A., C. Turpin-Petrosino and S. Guckenburg (2010) *Formal System Processing of Juveniles: effects on delinquency*, Campbell Systematic Reviews, 2010: 1

Pratt, T. and F. Cullen (2000) 'The empirical status of Gottfredson and Hirschi's general theory of crime: a meta-analysis', *Criminology*, 38 (3), pp.931-64

Pratt, T. and F. Cullen (2005) 'Assessing macro-level predictors and theories of crime: a meta-analysis', in M. Tonry (ed.), *Crime and Justice: an annual review of research*, Chicago: University of Chicago Press

Te Kawa, D. and K. Guerin (2012) 'Provoking debate and learning lessons: it is early days, but what does the Performance Improvement Framework challenge us to think about?, *Policy Quarterly*, 8 (4), pp.28-36

Tew, J., L. Harkins and L. Dixon (2013) 'What works in reducing violent re-offending in psychopathic offenders?', in L. Craig, L. Dixon and T. Gannon (eds), *What Works in Offender Rehabilitation: an evidence-based approach to assessment and treatment*, Chichester: Wiley

Weisburd, D. and A. Piquero (2008) 'How well do criminologists explain crime? Statistical modelling in published studies', in M. Tonry (ed.), *Crime and Justice: an annual review of research*, Chicago: University of Chicago Press

Weisburd, D., C. Telep and A Braga (2010) The Importance of Place in Policing: empirical evidence and policy recommendations. Stockholm: Swedish National Council for Crime Prevention