Children and Youth Services Review 33 (2011) 2076–2086

Contents lists available at ScienceDirect

Children and Youth Services Review

journal homepage: www.elsevier.com/locate/childyouth

Reframing child protection: A response to a constant crisis of confidence in child protection ☆

James Mansell *, Rissa Ota, Ricus Erasmus, Kip Marks

Child, Youth and Family, Ministry of Social Development, Bowen State House, Bowen Street, Wellington, New Zealand

ARTICLE INFO

Article history:
Received 17 January 2011
Received in revised form 9 April 2011
Accepted 11 April 2011
Available online 15 April 2011

Keywords:
Child protection
Reform
Performance
Decision errors
Child deaths
Signal detection

ABSTRACT

Child protection systems appear to be in a continual crisis of confidence. They get criticised for not doing enough to protect some children, whilst at the same time being criticised for being too intrusive or not managing demand. This constant balancing act drives almost continual reforms, none of which appear to reduce further crises of confidence. The central issue facing tertiary child protection systems stems from their function as makers of sometimes highly uncertain risk screening decisions. Uncertainty leads to errors; false positives and false negatives. Two recurring issues challenging child protection agencies are concerns about these errors. Fears about doing too much are concerns about false positives and fears of doing too little are concerns about false negatives. The need to address both issues within the context of uncertain high stakes decision making, in a highly risk intolerant environment leads to poorly formed sentinel event driven policy that in turn leads to organisational fragility. A decision outcome-based performance model based on Signal Detection Theory provides indicators that explicitly outline the link between these two strategic issues facing child protection systems. This has improved dialogue, understanding and support from sponsors. It has led an informed focus on improving decision making and stabilisation of decision thresholds. It demonstrates that Child protection systems are in fact very responsive and do perform well in their decision making (risk screening) function. Social work decision makers provide value in their decision making in spite of highly uncertain decisions to make. Child protection systems do not need reform, they need to be "reframed" to better understand true performance and so avoid poorly informed reactive policy responses to the genuine challenges that they face.

1. Introduction

Confidence in child protection systems across the United States, Australia, the United Kingdom, Canada and New Zealand appears to continually be in a state of a crisis. The types of crises experienced within these countries are remarkably similar. Criticism stems from a seeming inability to stop the deaths of children due to maltreatment. At other times child protection systems are criticised for doing too much: "Families who are subject to the unparalleled number of unsubstantiated reports are forgotten casualties of systems that focus on endless investigations…" (Lonne, Parton, Thomson, & Harries, 2008 p 9). Child protection services have become so 'forensic' and focused on detecting new cases of child maltreatment that capacity to provide adequate services is diminished and demand then overwhelms caring services. Constant injections of funding, resources and attempts at reform have consistently failed to avoid further crises. Not surprisingly staff morale is low, and turn-over high.

Child protection commentators and reformers generally agree that there is an intractable problem facing tertiary child protection systems (Lonne et al., 2008). However, they tend to disagree on the causes of systemic failure and on what can and should be done about it.

Although somewhat of a simplification, most calls for reform can usefully be described as either an attempt to (1) increase the level of surveillance and response, or to (2) decrease powers and mitigate overly excessive responses. Other reformers are more conflicted in their views, seeing as they do the need for reforms that will provide more responsive services to avoid further child deaths *and* also reforms to reduce a focus on less vulnerable cases. Some folk go as far as to assume systemic failure within protection systems and so seek system wide reform.

1.1. Reforms aiming to 'Do more'

High profile child deaths due to maltreatment drive recommendations to 'do more', and ensure such tragedies are avoided in the future. Thus child death reviews, the media, governments, and child protection experts recommend greater interagency collaboration and

☆ The views expressed by the authors do not necessarily reflect those of their agency or of the Government. They are part of a free and frank conversation and ongoing research within this area.

* Corresponding author.
E-mail addresses: james.mansell006@msd.govt.nz, jhmansell@gmail.com (J. Mansell), rissa.ota001@msd.govt.nz (R. Ota), ricus.erasmus001@msd.govt.nz (R. Erasmus), kip.marks007@msd.govt.nz (K. Marks).

0190-7409/$ – see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.childyouth.2011.04.019

information sharing, increased responsiveness to reported concerns, and sometimes even mandatory reporting or universal risk screening of all children. These kinds of recommendations often drive organisational system changes such as the introduction of contact centres, and an imperative for increased documentation or risk management practices to ensure each notification gets assessed in time. Practice imperatives shift towards capturing all child maltreatment concerns even when there are doubts. The assumption is that increased surveillance, information and action, and increased budgets, will mean that fewer tragedies occur in the future (Mansell, 2006a, 2006b).

1.2. 'Do less' reforms

Another reform agenda is for tertiary child protection systems to 'do less'. When this reform is driven by practitioners it tends to be driven by concerns that child protection systems are becoming overly forensic and paternalistic, and can damage children and families. There is concern that an unbalanced focus on investigating notifications rather than caring for children already in our care misdirects resources. This leads to poorer outcomes for the children we do know are at risk of child maltreatment. "In the process of detecting true incidents of abuse, large numbers of families suffer the distress of investigation with no help being offered." (Munro, 1999b, p 120). Worries are that highly coercive powers to separate families can be undertaken with little or no consultation, can lead to worse outcomes, and can target and over represent marginalised communities such as indigenous populations (Lonne et al., 2008). The reforms typically suggested here are characterised by attempting to halt the increasing numbers of investigations or removal of children into care.

> "Too many reports are being made to DoCS [Department of Community Services, New South Wales] which do not warrant the exercise of its considerable statutory powers. As a result, much effort and cost is expended in managing these reports, as a result of which the children and young people the subject of them receive little in the way of subsequent assistance, while others who do require attention from DoCS may have their cases closed because of competing demands on the system (that is, insufficient resources)." (Wood, 2008, p III).

From a decision making perspective these kinds of 'do less' concerns have led to significant reforms such as Family Group Conferences in New Zealand, and differential response models in the United States, Australia and New Zealand. All are attempts to triage notifier concerns away from the highly coercive powers of statutory protection agencies. The range of 'do less' policy positions attempt to balance child protection with other social responses. Proposals in Australia include 'Inverting the Pyramid' to provide families with better access to support options that meet their needs, before they reach crisis point (ARACY, 2008), and moving to a second tier response to the issue of child maltreatment via a 'public health model' (Scott, 2006b) or even prefer investment in primary community strengths to prevent child maltreatment. An example of the later is Melton's neighbourhood-based strategy to create "Strong Communities for Children" (Melton & Barry, 1994).

A more mundane driver of 'do less' criticism comes from treasury officials who become concerned about increased levels of 'soft demand' investigations with no findings. Their concerns being that continual increases in budgets fail to demonstrate corresponding improved value. Within protection organisations front line social workers struggling under increased workloads also want to formally or informally 'manage demand', either by increasing capacity (asking for more money) or raising the threshold for entry. Sometimes for 'do more' reforms there is confusion or obfuscation about the main driver of introducing such reforms as differential response models that aim to manage demand (cost and capacity) as well as in the belief that child protection systems are often not the best place to forward all types of concerns in the first instance (Mansell, 2006b).

Research in child protection systems in other child welfare jurisdictions have been criticised for drawing too many cases into the child protection net, only to drop them off at a later stage, with no services or support being provided (Department of Health (UK) 1995). Early classification at the reporting stage of a 'child abuse' tends to stick with cases being labelled 'child protection' until the investigation is completed (Thorpe 1994).

Current preoccupations with risk and risk assessment seem not entirely motivated by concerns about the safety and wellbeing of children. They are just as likely to be driven by the need to provide mechanisms to ration resources and allocate responsibility and subsequently blame (usually to social workers) when true positives are missed and cases 'go wrong'. Short term risk assessment and investigation are dominating activity, with correspondingly fewer resources or services being provided to the most vulnerable children over time. Child protection systems are putting large numbers of families through an investigation process and, in doing so, are losing sight of children's needs.

That reformers have been placed into two camps is somewhat of a simplification. Interested and thoughtful persons are likely to be more conflicted in their views. Many will be wishing at the same time to become more responsive to avoid further child deaths, whilst also not wanting to waste resources on lower needs cases. However, the straw-man is accurate in so far as each particular kind of reform can be described as an attempt to (1) increase the level of surveillance and response or (2) decrease, control or mitigate excessive responses.

We argue shortly that these two quixotically opposing goals are in fact tilting at the same windmill.

1.3. Systemic failure?

Concerns about both kinds of failures (not doing enough or doing too much) can drive a third breed of reform agenda. Reformers may note that such a systematic and ongoing crisis of confidence must be due to systemic organisational failure and so need system wide reform. Explanations as to the cause of system failure vary.

The perspective articulated by Lonne et al. (2008) attributes child protection system failures as a result of social work practice moving away from a professional responsibility model. They are a good example of practitioner reformers who suggest that systemic failures are as a result of rise of 'neo-liberal' agendas and 'managerialism' that do not understand or build systems that suit social work practice. Conversely, the persons they criticise ('managerialist' 'neoliberals') imply the opposite, that protection agencies are being poorly run, need better processes and have ill defined goals and poor practice (Ministry of Social Development & Treasury, 2003).

Polemics aside, it would be fair to say that the one thing these reformers share is the frustration that none of their reforms appear to be successful as new crises of confidence emerge after the introduction of their respective reforms. 20–30 years of attempts on both sides has not stemmed the tide of criticism as reforms appear to continually fail or unravel.

The thesis presented here is that the 'do more' and 'do less' reformers and the system wide reformers have misunderstood the problem. Our thesis is even stronger than that, in some cases we think that these kinds of reformers are the problem and not the solution.

We argue here that there are in fact three orders of challenge facing child protection systems. Firstly, managing 'wicked decisions', secondly, the reactive responses or well meaning reformers to perceived failures, and thirdly, organisational fragility through constant deep system instability.

1.4. First challenge: wicked decisions

The first order challenge is in managing decision making under high levels of uncertainty. For example, removal decisions or whether-to-investigate decisions.

Munro (1999b, 2003) recognizes that child protection systems have the challenge to make high impact decisions often in conditions of a high-level uncertainty. It is well recognised that the intake decision is made sometimes under intense time pressure and with minimal information that is often conflicting and uncertain (Benbenishty & Chen, 2003; Child, Youth and Family, 2000; Gambrill, 2005; Gambrill & Shlonsky, 2000; Mansell, 2006a; Munro, 1999a). When a potential concern is forwarded, this uncertainty drives a difficult balancing act between managing the decision risks of either doing too little (false negative) or doing too much (false positive).

There can be high-stakes consequences for children, their families, social workers, the agency and notifiers involved if there is an error in decision making, either where a child with statutory level concerns is not escalated (a false negative) or where a child without statutory level need is escalated (a false positive) (Mansell, 2006a; Munro, 2004; Scott, 2006a). For instance, a false negative decision can result in further maltreatment for the child. A false positive decision can lead to the unnecessary investigation of a well-functioning family. This action may damage relationships within already vulnerable families and/or wasting resources that would otherwise be better spent on known clients with high needs; so indirectly harming children already in statutory care.

Reformers are understandably worried about the consequences of errors. However, their reforms treat false positives and false negatives separately. What they fail to appreciate is that these two issues are intrinsically linked and any single issue lead reforms aimed at reducing one kind of error will unavoidably have the unintended consequence of making the other kind of error more likely.

Uncertain decisions become 'wicked decisions' when they are made by public sector officials, can have high impact harmful outcomes, are made within a high transparency environment and in front of a highly risk intolerant public. These decisions are 'wicked' in the sense of it being difficult to resist over reacting to one or other kind of error. The first order challenge is the genuine problem of how to do the business of making 'wicked' decisions every day and making errors every day in a manner that mitigates the risk of giving rise to the second order challenge — poor responses to perceived decision failures.

1.5. Second order issue: responding well to errors

The making of difficult decisions is not the cause of the crisis of confidence in child protection systems. The offspring of uncertain decisions are errors, and these are common and unavoidable. However, when high impact becomes high profile then this can lead to the formulation of reactive crisis driven policy. Responses to errors are typically single-issue focused and so apt to unravel over time as their unintended consequences take over:

> "Moving the threshold to reduce one type of error automatically increases the other type. Applying this to child protection history, when society was outraged by the death of Maria Colwell [United Kingdom 1973], professionals responded by lowering the threshold for intervention to minimise the chances of missing another child and such extreme danger. This necessarily led to more families with low actual levels of abuse being caught up in the net … [Fearfulness of missing a serious case] … had the inevitable result of making professionals prefer to overestimate the risk of abuse since the consequences of underestimation were so severe. This in turn provoked a backlash, in with critics claiming that social workers were overreacting and interfering in family life. Meyers (1994 p 4) gives examples of the emotive language used. The child protection system was described as "trampling the rights of innocent citizens" and engaging in "hysterical witch-hunts"; social workers were likened to "Nazis", McCarthyite persecutors and the KGB." (Munro, 1999b, p 119)

The objectives of one-sided 'demand management' or 'risk management' reforms ultimately unravel. The unintended consequence of reactive reform about one kind of error is to unravel the previous round of reforms about the other kind of error and so on.

Another problem with single error driven reform is asymmetric accountability. Attention focuses mostly on false negatives since the impact of a child dying is much more graphic and obvious than the less tangible and visible impact of lost time spent on caring for children we already know to be at significant risk. This asymmetric accountability in favour of focusing on false negatives is the driver of lop-sided reforms over time tending to favour ever more powers of coercion, forensic investigation, and financial muscle to social workers to act faster and with less constraint to avoid missing the opportunity to act to save children (false negatives). So Family Group Conference decision making can get shunted to the back of several layers of agency investigations decisions. Funding is redirected to tertiary child protection responses and away from other targeted, non-coercive services in the effort towards zero tolerance to false negatives.

The second order issue facing child protection systems is poorly conceived policy or system responses to the symptoms of the first order challenge (errors). The bane of child protection systems are persons (internal and external to agencies) who drive policy and practice off single case studies evidence.

1.6. Third challenge: organisational fragility through constant deep system instability

The offspring of reformers' poorly informed responses to these kinds of errors is deep system instability via almost constant, ineffective and reactive, single-issue policy and system changes. The system tends to oscillate over time constantly between responding to imperatives to 'do more' or 'do less', often at the same time (Mansell, 2006a, 2006b).

Constant change erodes morale and confidence. Constant criticism leads to a defensive 'organisational risk management' stance. This reduces the ability to learn from our past mistakes. It erodes the ability to have a constructive dialogue where errors are admitted to be common and unavoidable. The system and its processes become increasingly 'over fitted' due to error led design. Constant change, the threat of a poorly judged response to social work and working within a highly pressured job with minimal trust and a complex process and paperwork obsessed system inevitably and understandably leads to high turn-over. High turn-over in turn leads to diminished levels of experience at the front line and within management. This lack of depth further compromises performance.

In summary, the crises of confidence emerge from a misguided focus on the proximate symptoms of decision making under uncertainty — the necessary two kinds of errors. Judging the performance of child protection systems by a piecemeal focus on one kind of error and on single cases of errors is a poor source of performance information. This poor understanding of performance leads to poorly considered reform where the unintended consequences of attempts to reduce one kind of error unravels reforms aimed at reducing another kind of error. This constant oscillating reform agenda leads to the third order issue facing child protection systems — organisational fragility through constant change of systems, processes and leaders (Mansell, 2006a).

It is unsurprising that child protection systems become organisationally fragile and moribund with procedural risk management and monitoring, yet still unable to demonstrate value or seem to avoid further crises of confidence as demand continues to spiral out of control and agencies still fail to save children from maltreatment.

The main lever for making protection agencies more resilient must therefore be to address directly the difficulties faced in the first order challenge. The challenge is not that errors are constantly made. It is that they are constantly misinterpreted or misused to drive single issue policy. The challenge is therefore enabling people to understand what *should* be done in response to errors. Not doing this inevitably leads to constant reforms and their constant failure. The first order challenge is real, how do you scaffold the system to allow for difficult decisions and their high impact constant flow of errors? How do you do so without falling into the trap of the self-manufactured second and third order issues?

1.7. Responding to the challenge to mitigate the risk of our system working against us

We argue in this paper the real challenge facing child protection is to balance and improve decision making. We argue this does not require *reforming* child protection systems. It requires *reframing* the core understanding of the role of child protection and what *should* and *can* be learnt from the inevitable errors; i.e. working to improve dialogue, understanding and support for the real challenge of making wicked decisions. Focusing on improving decision making is required for its own sake to make fewer errors of both kinds where possible. Reframing is also required in order to mitigate the risk of poorly conceived reactive crisis-driven policy that makes things even worse than they need to be so avoiding falling into the second and third order traps as much as possible.

Research in New Zealand into these dynamics underlying child protection systems (Mansell, 2006a, 2006b) led to acknowledging that risk screening decision making is a core service that needs to be measured, monitored and improved. By having a business model and performance framework that more accurately reflects the nature of the risk screening task of child protection systems it is possible to understand, demonstrate and improve our decision making performance. In short, the goal is to bridge the gap in understanding and dialogue between the 'do more' and 'do less' reformers.

This paper shares the new performance framework for one of our core services – risk screening decision making – by looking at the outcomes of these decisions. It directly measures the challenge facing front line decision makers. In doing so it escalates the challenges faced at the front line each day to make difficult decisions and allows this to be communicated to senior management and external sponsors in a meaningful way.

The trouble is not as Lonne suggests that performance-orientated managers wish to have information about the value of child protection systems. The problem is that the wrong kinds of process and symptom measures are collected. What is needed is good performance information about risk screening decision making under conditions of high uncertainty. This requires systematic information about all errors and non-errors, not merely single cases of error as tends to be the case in child death reviews.

The model developed here focuses on the 'intake' decision. One of the central tasks of tertiary child protection systems is deciding when concerns forwarded by notifiers are sufficient to escalate to a statutory investigation (or not). Specifically, the task is to try to escalate children with statutory level need, whilst at the same time leaving alone (or sending elsewhere) children with other types of need.

The performance of a child protection agency with regard to its ability to detect and escalate the right types of cases, where children are at risk, can be usefully modelled using some of the methods and principles of Signal Detection Theory. Conceptually Signal Detection Theory accurately reflects the generic performance considerations of the practice of making social work intake decisions. Social workers have to sift through a lot of information and misinformation or irrelevant information to determine who should be escalated towards a statutory response and who should be sent elsewhere (or given advice, etc). Standard measures used in Signal Detection Theory provide a simple and easily understandable set of performance indicators for child protection decisions.

2. Method

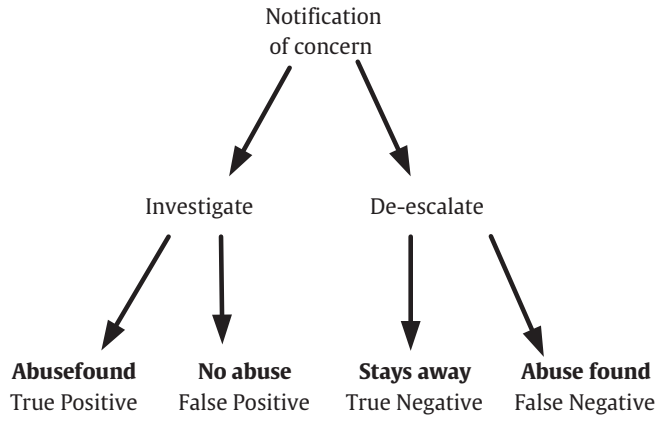
The model presented here is based on interpolating estimates of decision outcomes and using this information to develop several other core metrics that enable a signal detection model to be built and used to understand the performance of intake decision makers.

At a very simple level, there are two responses for an intake decision. A social worker can decide to escalate a notifier concern towards a statutory level response (such as a statutory investigation) or they can de-escalate it (away from a statutory investigation). In New Zealand the decision to escalate towards an investigation is often called Further Action Required (FAR). De-escalation may include No Further Action (NFA), referral to other services or Adding Notes to a case file without undertaking any further case work, or send notifications down a non-statutory "differential" response.

Note that an alternative decision metric is available using the urgency rating of which there are five levels in New Zealand: No further Action, Low Urgent, Urgent, Very Urgent and Critical. This was found to provide finer grained information on decision performance. The principles of the 5-way model decompose to the 2-way model. Results from the 2-level (escalation decision) and the 5-level (urgency decision) methods are both provided below.

The escalation rate (using the 2-level model) is the tendency to escalate concerns towards a statutory level investigation. It is the Further Action Required (FAR) rate, which is the proportion of engagements that are FARed.

For the measures developed here, the outcome of the decision to escalate or not is estimated from the presence of "Findings" (of maltreatment or behavioural difficulties) and/or "Outcomes" (decisions to do more statutory work) subsequent to an investigation. Any Finding or Outcome within one year from the intake decision in question is assumed to reflect a statutory level child protection need/risk. Escalation decisions where there are no findings and no outcomes within one year are assumed to indicate a client with lower needs than require a statutory level response.



The one year reference period was chosen after checking alternative definitions (6 months to 3 years) and was found to be the best trade-off between availability of the information and data censoring (censored in the sense of not knowing that a person will re-engage with a finding of maltreatment after the reference period). This indicator information can be used to place each engagement decision into one of four high-level decision outcomes:

- True positive (TP): A client was escalated to investigation and there was some form of maltreatment or a need requiring a statutory level response recorded within a year.

- False positive (FP): A client was escalated to investigation and there was no maltreatment or a need requiring a statutory level response recorded within a year recorded within a year.
- True negative (TN): where a client is not escalated and there was no maltreatment or a need requiring a statutory level response recorded within a year recorded within one year. They "stay away".
- False negative (FN): A client is not escalated and there was maltreatment or a need requiring a statutory level response recorded within a year recorded within one year.

Note that for any particular case these assumptions may be incorrect. For example, a young person may have been notified and have not been at risk at the time the intake decision was made. In this case the decision maker correctly does not escalate. Three months later an entirely new risk is introduced into the family context and they are the subject of another notification and there is a finding of maltreatment made within a year. In this case the first decision would have been a true negative but is misclassified as a false negative. Notwithstanding the potential for some cases to be misclassified, at an aggregate level the indicators are found to be accurate enough to usefully capture the core decision making dynamics.

Note also that a 'false positive' does not indicate that *no* needs are present. For example, a family with employment and alcohol issues may rightly be considered not appropriate for a statutory child protection response since the children are safe and loved. The decision to undertake a statutory investigation was 'false' in relation to the need for statutory response but not false in the sense that there are no needs or that no other form of support would be welcome to achieve a better outcome for this family. The judgment of decision accuracy is always in relation to the purpose of the decision in question. For Intake decisions this is making a decision about exercising considerable statutory level powers of investigation. The decision is not being judged on its ability to detect *any* form of need. Indicators could be designed that would measure this different aim if required. Using the example above, a different measure of resource targeting decision performance for a differential response might have considered an escalation down a pathway to have been a 'true' positive, though 'false positive' for the purposes of commencing a statutory investigation.

The decision to investigate requires an indicator of statutory level need (this being the assumed goal of intake decisions considered here). In this case the assumption is that a social work investigation finding and decision to do further work (in Child Youth and Family called an 'Outcome') both indicate an identified need requiring a statutory response. There is no infallible indicator for "statutory level needs" so the assumption taken for these measures was to use the investigating social worker's clinical judgment on whether there were statutory level needs or not. Better indicators may be developed in the future where needs assessments are more fine grained and have higher inter-rator reliability (greater consistency between "rators") in arriving at substantiation decisions.

An additional limitation is that, where cases are de-escalated (not investigated), and do not get re-notified, there may well be a missed cases of child maltreatment that are incorrectly considered 'true negatives'. Due to the way the system works this information is censored. This means that some 'false negatives' will be incorrectly assigned as 'true negatives'. Therefore the true level of 'false negatives' is always under-reported.

The limitations for any single case should be carefully considered. However, the resulting measures do portray the underlying dynamics important to overall decision making practice and so have been useful in New Zealand to provide some, albeit not perfect, information. Importantly they have allowed the relationship between core strategic issues facing child protection systems to be demonstrated quantitatively for the first time and so better used to inform planning and dialogue.

The four decision outcomes are used to make the two further Signal Detection Theory indicators; Sensitivity (Se) and Specificity (Sp).

Sensitivity (Se) is a measure of the sensitivity of the intake screening processes to detecting children with *statutory level concerns*. Sensitivity answers the question *How good is the agency at detecting statutory level child maltreatment*? Scores closer to one reflect a better ability to detect maltreatment hazards for children. Sensitivity is an indicator of an intake decision maker's ability to escalate cases of child maltreatment and avoid missing cases of child maltreatment ('false negatives'). For example, if there were 200 persons at risk and 180 were escalated then sensitivity would be 90% and 10% would have been 'false negatives'.

$$Sensitivity(Se) = \frac{TP}{TP + FN} \quad Se = \frac{180}{180 + 20} = \frac{180}{200} = 0.90$$

Specificity (Sp) is the proportion of persons who would not have had a statutory finding who were correctly not escalated. The higher the specificity, the better the agency is at avoiding 'false positives' (sometimes called "soft demand"). For example, to carry on with the example above, if there were an additional 1000 concerns notified that were not at high risk but 450 get escalated for an investigation then specificity would be 65%. The portion of non-maltreatment cases *not* being escalated.

$$Specificity(Sp) = \frac{TN}{TN + FP} \quad Sp = \frac{650}{650 + 450} = \frac{650}{1000} = 0.65$$

In practice the complement of specificity, '1-Specificity', is used since this better reflects the desire to avoid false positives. Using the example above, the 1-Sp would have been 45%; i.e. 45% of the persons who are not at risk were nevertheless escalated.

$$1-Sp = \frac{FP}{TN + FP} = \frac{450}{1000} = 0.45$$

2.1. Receiver Operating Characteristics ROC curve

Sensitivity and specificity provide important information about the strategic imperatives to detect risk and avoid investigating families where this is not necessary. Understanding the relationship between sensitivity and specificity provides additional information about the level of uncertainty of the risk screening trade-off. In Signal Detection Theory this is traditionally called a Receiver Operating Characteristics (ROC) curve (Fawcett, 2005). ROC curves traditionally use Sensitivity on the Y-axis and complement Specificity (1-Sp) on the X-axis.

3. Results

The indicators of decision outcomes (Fig. 1) are able to be estimated and tracked over time using New Zealand data from Child, Youth and Family.

The complex interaction between these four decision outcomes can then be reduced to two main indicators (Sensitivity and Specificity) that target the main strategic issues facing child protection systems. Sensitivity and Specificity clearly have a relationship between each other. They both move in the opposite directions at the same time (i.e. as one goes up the other goes down).

There is also a strong correlation between Sensitivity, Specificity and the escalation rate of lodged notifications — in New Zealand called the 'Further Action Required' or FAR rate; portion of reports of potential maltreatment – notifications – that were escalated that month.

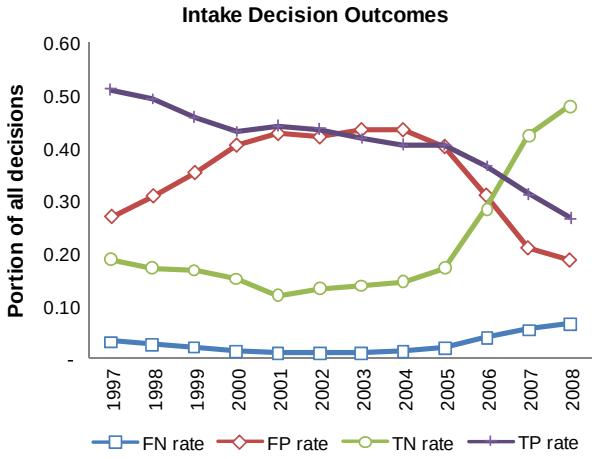


Fig. 1. Intake decision outcomes.

When the FAR rate changes so does the mix of decision outcomes (the level of Sensitivity and Specificity) as is observed by the inflection points for Sensitivity, Specificity and FAR rate all change direction at the same time. See Figs. 2 and 3. The relationship between the intervention threshold (FAR rate) and the level of Sensitivity or Specificity able to be achieved can be observed more clearly in the scatter plots (Figs. 4 and 5).

This shows that as the agency errs towards escalating a high proportion of notifications (FAR = 90%) sensitivity to risk improves (Se = 98%). However, this comes at the *unavoidable* cost of a decrease in performance of avoiding false positives (Sp < 20%). This meant the process of focusing resources on finding more cases of child maltreatment came at a trade-off in misallocation resources by investigating 'false positives' (low or minimal needs families).

Alternatively, an agency erring towards improving specificity (FAR = 70%) leads to a reduction in soft demand (Sp > 40%) and also leads to a higher chance of missing more high-risk cases (Se < 90%).

Whilst child protection agencies may introduce reforms and policy settings that aim to manage the balance of errors (one way or the other), they do not have the choice to avoid all errors. Errors of some kind will *always* be made.

4. Receiver Operating Characteristics (ROC) curve

Points on a ROC curve depict the effect of decision making performance and intervention threshold together. Each point on the graph below represents the decision performance result for a year. The graph's points start at 1997 in the middle, move to the right by 2001 then shift back towards the left until 2008.

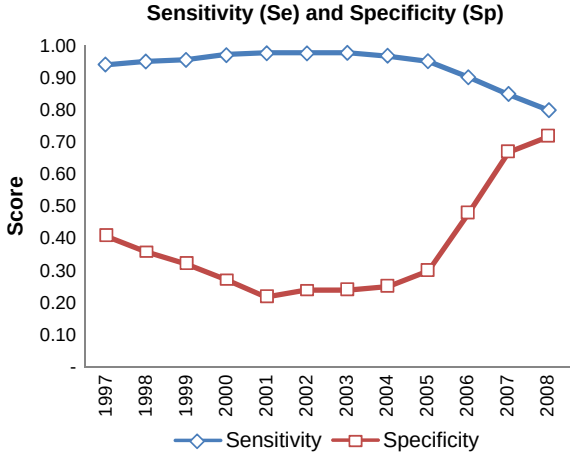


Fig. 2. Sensitivity (Se) and Specificity (Sp).

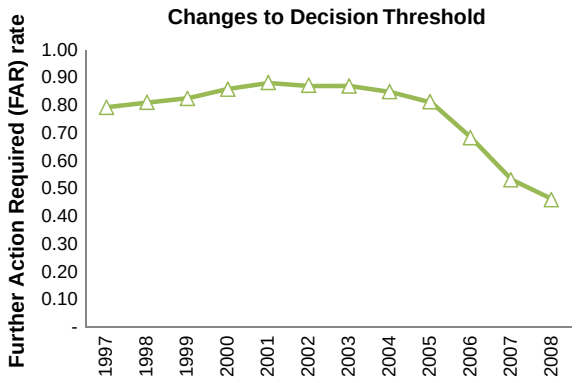
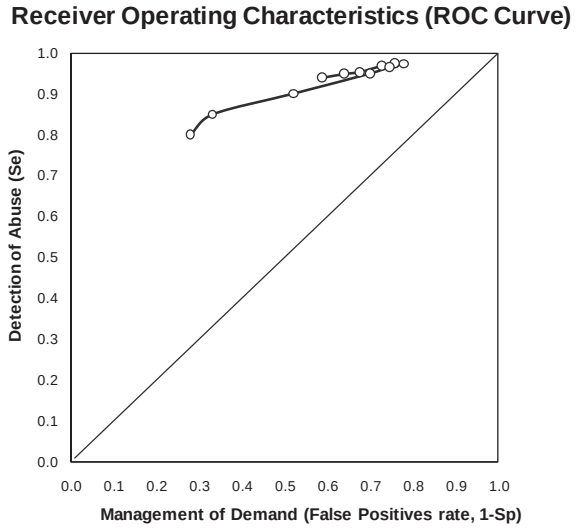


Fig. 3. Changes to decision threshold.

The results below summarise the level of uncertainty facing intake decision makers and the level of value they add through their decision making. New Zealand intake decision makers are better than random decision makers since the decision points are all well above the 45 degree line. Therefore intake decision makers are demonstrably adding value in targeting services and perform better than chance. They are by no means perfect and do not occupy the top left hand corner. Therefore errors of both kinds are still common.



Oscillating behaviour can be clearly observed in the changing intervention thresholds (FAR rate) and how this leads to oscillations on the ROC curve. Three distinct periods of intake risk screening behaviour can be observed using the inflection points for the escalation threshold, Sensitivity and Specificity. Looking at average decision outcomes for these three periods clearly shows the effect on the mix of errors made during these periods (Table 1).

4.1. Branch-based demand management in the late 1990s

In the 1990s, under a branch-based model of intake (prior to the introduction of a centralised contact centre) the escalation (FAR) rate was 80%, Specificity was 37% and sensitivity was 96%. This is consistent with the policy settings at the time. Several reports, including the Duffy review (2000), Mick Brown report (2000), and Weeks review (1994), had concluded that child protection service

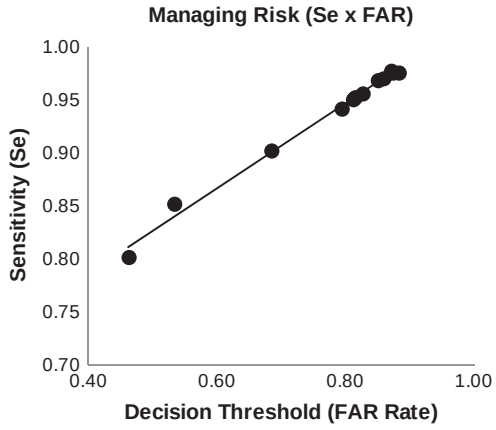


Fig. 4. Managing risk (Sex × FAR).

in New Zealand was struggling with demand and referring a greater proportion of cases elsewhere. Social workers were observed to be managing demand via a 'workload bias' when making intake decisions (Duffy, 2000). Concern about the threshold for investigating was one of the drivers of introducing a national contact centre to provide greater access to services (Duffy, 2000; Mansell, 2006a).

> "This move towards providing a Call Centre was aimed at separating intake prioritisation decisions based on the criticality of cases, from resource allocation decisions, to provide an enhanced level of service." (Department of Social Welfare, 1999)

Note also that notifiers were similarly criticised by reformers concerned that they were not doing enough to identify child maltreatment.

4.2. Highly precautionary period from 2000 to 2004

During the early 2000s, in response to the 'workload bias', increased concerns over family violence, and several high-profile child deaths the threshold and system itself shifted. Child, Youth and Family was asked to "escalate when in doubt", "lodge every concern" and "lodge every sibling". Notifiers were asked to work more closely with Child, Youth and Family and criticised for not notifying (Mansell, 2006a). The reform during this period aimed to become more responsive and so the escalation rate increased. The probability of being escalated for investigation rose to 85% by 2001. This

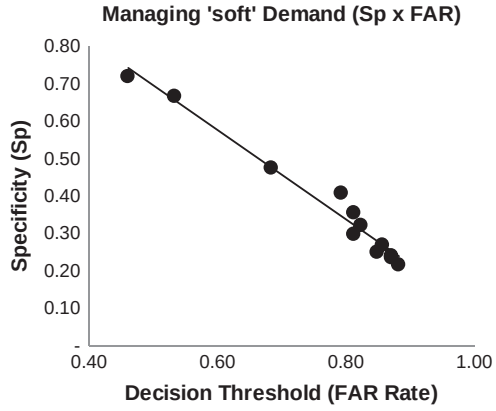


Fig. 5. Managing 'soft' demand (sp × FAR).

Table 1
The effect on decision outcomes of oscillating thresholds in Child Protection.

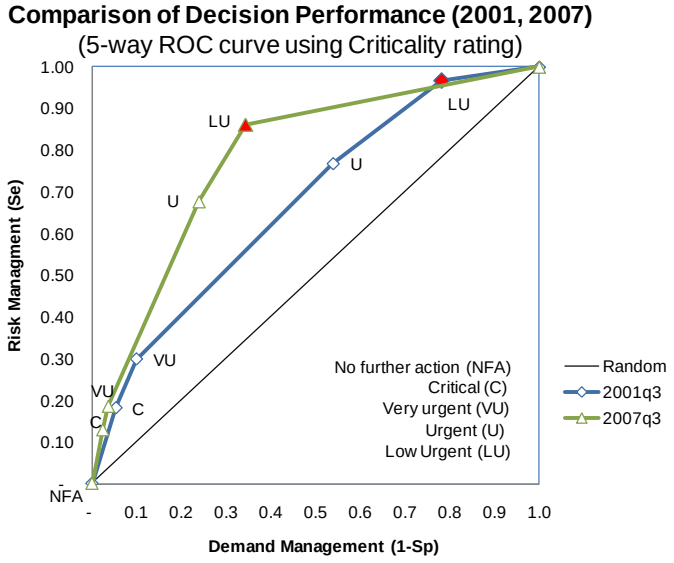
1997–99	Findings	No findings	Total	
FAR	49	31	80	Se = 0.96
Not FAR	2	18	20	Sp = 0.37
	51	49	100	
2000–04				
FAR	42	43	85	Se = 0.98
Not FAR	1	14	15	Sp = 0.25
	43	57	100	
2005–07				
FAR	35	29	64	Se = 0.90
Not FAR	4	32	36	Sp = 0.52
	39	61	100	

translated into improved levels of Sensitivity (Se = 98%). The probability of missing children at risk ('False Negatives') halved from 4 per 100 decisions to 2 per 100 decisions. These reforms achieved their objective to miss less cases of child maltreatment. However these improvements came at the cost of a large decrease in Specificity from 37% to 25%. 'False positive' decision errors increased from 31/100 to 43/100 (39% increase). This reflects the necessary trade-off in risking the misdirection of resources ('false positives') to improve sensitivity to risk of missing cases ('false negatives'). The effect of doing so is less resources to attend to young persons who are already known to be at risk. I.e. who are already within the system.

4.3. Multi-agency demand management (2005–2007)

Since 2005 due to the combined effect of increased responsiveness of notifiers (leading to three times as many notifications), and agency increased responsiveness (leading to twice as many investigations) the imperative was to manage demand. This was necessary to remain within funding constraints and to manage capacity-constrained social work resources to ensure that some resources remain available to care for existing clients. Where demand outstrips capacity then either capacity needs to be increased or demand slowed down. The only way to reduce demand is to raise the threshold before notifications get escalated to investigation or attempt to change notifier behaviour. A differential response model began in 2004. Work was undertaken to improve the quality of notifications perceived as low value (lots of false positives) through improving the Police "POL400" form sent to Child, Youth and Family. In addition the protection agency escalation rate was decreased (FAR rate = 68%) to manage demand. This translated into a huge increase in Specificity (from 25% to 52%, an increase of 108%). The aims of reformers concerned to manage demand and stop becoming overly 'forensic' were successful at reducing the number of false positives (unnecessary investigations) from 43/100 to 29/100 (−33%). However, these demand management reforms also had the unintended consequence that this led to decreased sensitivity to child maltreatment. Sensitivity decreased from 0.98% to 0.90% (−8%). The number of children who were NFAed and returned with findings (false negatives) increased from 1 percentage point during 2000–2004 to 4 percentage points in 2005–2007.

In the 5-way ROC curve below the points closest to the bottom left corner are the 'Critical' urgency rating ('C'). Dots closest to the top right corner are the 'Low Urgent' ('LU') rating (in filled markers). The 'Low Urgency' rating is equivalent to the 2-way threshold for escalation. Points below this relate to the urgency with which escalated cases were assigned.



The changes in threshold between 2001 and 2007 occurred for the decision outcomes for all criticality ratings: Low-Urgent (LU), Urgent (U), Very Urgent (VU) and Critical (C). The ROC curve itself is steeper and so indicates better overall decision making performance. In 2007 the dramatic improvement in demand management (1-Sp) is bought at a much smaller decrease in Sensitivity to risk than it would have in the past; i.e. less errors over-all.

The combined set of new intake decision performance indicators for the 2-way decision model is presented below (see Table 2).

The base number is the number of true positives plus false negatives. The base rate is the portion of decisions where an identified need requires some form of statutory response (as defined by front line social workers). The base rate has been declining steadily as the number of decisions to make (i.e. forwarded concerns by notifiers) has increased.

5. Discussion

The new indicators conform to expectations with regard to being able to monitor and observe the core dynamics around intake risk screening — in particular the trade-off between managing risk (not missing cases) and managing demand (false positives). The ROC curve also illustrates the level of uncertainty in making decisions — that there are a constant flow of errors. This is what would be expected if social work intake decision making was akin to signal detection — listening through a lot of noise to try to detect a true signal of maltreatment. It is an important step in communicating the link between 'doing more' or 'doing less' that has plagued child protection agencies.

There is a strong relationship between the escalation (FAR) rate and the outcomes of intake decision making. A high FAR rate leads to a greater proportion of soft demand (false positives) being escalated for investigation to find more maltreatment. Lower escalation rates, the tendency to err in favour of increased 'No Further Action' (NFA)) leads to better management of 'soft demand' (less 'false positives') but also results in a greater rate of 'false negatives' maltreatment where cases are missed the first time around. This trade-off between managing the risk of missing cases (false negatives) and managing constrained caring resources (by avoiding false positives) can be observed and demonstrated using a ROC curve graph. This also provides an indication of the level of uncertainty that social workers face when making decisions on the frontline.

The indicators all appear to be remarkably stable and provide consistent results over different time periods and different regions. This means that definitive trends can be observed. The results of this initial analysis are consistent with Signal Detection Theory and with the expected behaviour of the system (due to changes to policy settings), and is consistent with previous work on the drivers of demand — in particular the system dynamics work in 2006 predicted in advance that the system would oscillate towards demand management (increasing Specificity) after a period of focus on reducing the risk of missing cases (aiming for high Sensitivity to maltreatment) (Mansell, 2006a, 2006b).

For these reasons it can be reasonably expected that the indicators are providing genuine information about the main system dynamics driving the outcomes of intake risk screening decisions.

Some limitations are that: (1) waiting for investigations to be finished leads to delays in getting up-to-date information; (2) the true number of false negatives in the community will always be under-counted; and (3) reliance of substantiation practice to confirm whether a case requires a statutory response is not always reliable.

Notwithstanding these three limitations, the model does provide more practice focused and relevant information about a core service provided by protection agencies. It does demonstrate performance and sensitizes child protection agencies to the strategic issues driving policy changes and can track the influence of these changes on decision outcomes for children. In this sense this model is more informative than input orientated or process focused (queuing behaviour) performance measures that often have more distant relevance to social work practice or helping children in need. Most importantly the model provides more objective and reliable evidence about what decision making systems are working/not working.

The new measures illustrate that the 20 or so high profile child deaths are only a small fraction of the in terms of total number of errors made during the last 10 years. Errors of both kinds are a

Table 2
Decision making performance indicators applied to child protection intake decisions.

Eng start year	Intake decisions	Base rate	FAR rate	False negatives (FN)	False positives (FP)	True negatives (TN)	True positives (TP)	FN rate	FP rate	TN rate	TP rate	Sensitivity	Specificity	1-Sp	Area under ROC curve
1997	19,956	0.54	0.79	632	5384	3756	10,184	0.03	0.27	0.19	0.51	0.94	0.41	0.59	0.73
1998	22,447	0.52	0.81	572	6936	3884	11,055	0.03	0.31	0.17	0.49	0.95	0.36	0.64	0.72
1999	22,778	0.48	0.82	484	8043	3831	10,420	0.02	0.35	0.17	0.46	0.96	0.32	0.68	0.70
2000	21,634	0.44	0.86	287	8766	3269	9,312	0.01	0.41	0.15	0.43	0.97	0.27	0.73	0.71
2001	20,392	0.45	0.88	222	8,712	2447	9,011	0.01	0.43	0.12	0.44	0.98	0.22	0.78	0.68
2002	22,167	0.45	0.87	242	9326	2939	9,660	0.01	0.42	0.13	0.44	0.98	0.24	0.76	0.70
2003	25,937	0.43	0.87	253	11,261	3585	10,838	0.01	0.43	0.14	0.42	0.98	0.24	0.76	0.71
2004	32,470	0.42	0.85	430	14,098	4771	13,171	0.01	0.43	0.15	0.41	0.97	0.25	0.75	0.69
2005	40,749	0.43	0.81	836	16,400	7025	16,488	0.02	0.40	0.17	0.40	0.95	0.30	0.70	0.69
2006	50,966	0.41	0.68	2,008	15,811	14,513	18,634	0.04	0.31	0.28	0.37	0.90	0.48	0.52	0.72
2007	64,912	0.37	0.53	3,523	13,657	27,489	20,243	0.05	0.21	0.42	0.31	0.85	0.67	0.33	0.77
2008	78,771	0.33	0.46	5,178	14,747	37,832	21,014	0.07	0.19	0.48	0.27	0.80	0.72	0.28	0.78

common daily occurrence. Why wouldn't they be, this is exactly what uncertainty implies. However, the evidence based on single case studies and their profile in the media are what is currently used to inform policy and typically in crises situations. That is just bad science. These new measures look at *all* decisions and the range of outcomes (true positives, false positives, false negatives and true negatives). They do not asymmetrically focus solely on false negatives.

There is an unavoidable trade-off between making false positives (misallocating resources) and false negatives (missing cases) due to the high level of uncertainty inherent in making child protection intake escalation decisions. The child protection environment has a much greater public focus on false negatives (missing cases of maltreatment) due to the highly visible high stakes consequences these can have. This drives highly precautionary screening practice over time and this is true of New Zealand's child protection system. Child, Youth and Family has traditionally taken a highly precautionary approach to escalating cases (having an escalation rate of 81% per quarter between 1997 and 2007) to find children at risk — escalating when in doubt to not miss cases of maltreatment. The outcome of this is reflected in occupying the top right hand corner of the ROC curve achieving high Sensitivity (averaging 95% during the same period).

However, this precautionary screening to avoid missing cases of child maltreatment has lead to a high rate of 'false positives'. Thus Child, Youth and Family has been accused of not managing demand or diverting resources away from caring for children already in care.

Hyper-precautionary escalation for investigation is the product of wanting to not miss cases of maltreatment. But there is no free lunch in being hyper-precautionary about just this one side of the decision making challenge. Importantly the model now illustrates that the unintended consequence is to achieve a single extra true positive in 2000–2004 (a reduction in false negatives of one per 100) required an extra 12 false positive investigations. The cost in time, misdirected resources and harm can be borne into account. These results demonstrate that a large number of false positives are the product of a very precautionary approach to risk screening and so demonstrates the cost of a precautionary approach. The importance of this finding is that large numbers of false positives are *necessary* to attain a high level of sensitivity. So criticism about mismanagement of "soft demand" is ill founded.

Cost, time and expertise can be allocated to the detection of new hazards or dealing with the hazards found. To not balance the detecting task with the caring task is to trade off the needs of children at risk of undiscovered maltreatment with the needs of other groups of children and their families; (1) children already known to be at risk and needing the care services of the agency and (2) children who would be better assisted by non-statutory level response or better left alone entirely.

If sponsors wish to achieve a greater level of sensitivity to child maltreatment then the corresponding expected number of false positives can now be forecast accurately and be considered as one of the costs to achieve the required greater level of Sensitivity (Se).

To those politicians who hope to achieve 'zero tolerance' to another child death, think again. The ROC curve suggests a diminishing return on extra screening so that absolute elimination of 'false negatives' is impossible or comes at the cost of directing a huge level of resource into risk screening. Calls for mandatory reporting and universal screening fail to take into account the wholesale misallocation of resources towards forensic screening ('false positive' investigations) and the harms that this does to other children, their families and the community. There is no silver bullet or simple answer to the problem of making these highly uncertain 'wicked' decisions.

The new model and findings move past the unrealistic assumption and expectations that all risk (Se) and all soft demand (1-Sp) can both be managed independently of each other. The ability to indulge in one sided sentinel event driven policy is also mitigated to some extent when the true costs of doing so are transparent. Demand management initiatives will unravel or conflict with risk management initiatives. These two vexing strategic issues are two sides of the same coin. Both are products of the single process of making difficult and uncertain risk screening decisions.

The new measures outlined in this new model proivide a more nuanced view of performance, where the real goal is to stabilise the response threshold and improve overall decision making (reducing uncertainty), so reducing the level of trade-off between both kinds of errors (Mansell, 2006a, 2006b).

5.1. Oscillating trends

Notwithstanding the consistent tendency to emphasise risk management through a precautionary escalation strategy, as might be expected by a risk assurance agency like Child, Youth and Family, the threshold for intervention has oscillated during this period. The overall oscillating threshold behaviour is illustrated in the changes to the agency ROC curve and was predicted by research undertaken in 2005. This is a reflection of the changing reform imperatives in response to proximate crises driving towards both better manage risk and manage demand during different periods of time (Mansell, 2006a, 2006b).

The three observable periods with different escalation thresholds, driven by different concerns, lead to a different range of decision outcomes for children. The broader finding here is that intervention thresholds are responsive to policy settings from internal and external stakeholders and have lead to oscillation of the intervention/responsiveness threshold. Child protection agencies are responsive, to reformer's concerns … unfortunately.

Child, Youth and Family has often been criticised for poorly managing demand and as poorly managing the risk of missing cases. In fact the agency has demonstrably improved each of these issues — in turn. Child, Youth and Family has been exceptionally responsive to both kinds of concerns and admirably followed the reform agenda's of a host of external single issue focused experts.

As Munro points out, the real problem is not that child protection agencies are unresponsive or unable to perform. They are merely adapting in a "well-intentioned effort to respond to the unreal expectations of the general public" (Munro, 1999b p 125).

This also illustrates the disruption, unintended consequences and futility of continually running after one or other of two incommensurable goals; to manage all risk and all demand at the same time.

5.2. Learning how to make better decisions

Interestingly, it appears as though more balanced decision making in CYF since 2006 has lead to better decision making over all. Evidence from the 5-way decision model suggests that decision performance is better across all 5 level settings (all five points on the ROC curve) providing evidence that decision making has become more nuanced and considerate of the presenting case issues rather than defensively escalating nearly everything. The 2007 ROC curve is much steeper representing a better trade-off between each error type and so less errors overall.

The new indicators allow direct comparison between different intake decision making processes. This helps to clarify where decision outcomes reflect differences in thresholds and/or a difference in actual performance of child protection agencies.

Many of the differences between different intake decision processes in Child, Youth and Family have been threshold-based. At the national contact centre initial intake decision being made has always had a very high escalation (FAR) rate (90% plus) whereas the branch decision making has tended to have a lower FAR rate (as low as 40%) and as demand has increased this has lowered still further. The

overall performance (ROC curve) is similar for both notification channels. The largest difference is in the kind of errors made. The contact centre is much more likely to not miss a case of maltreatment (has a higher Sensitivity) and branches are less likely to escalate investigations with no findings (have a higher Specificity) and this is driven by different thresholds for escalation.

Policy and reforms in regards to management of intake aimed to introduce the Contact Centre to (1) improve the quality of intake decisions made and to (2) avoid a 'workload bias' and allow better access to investigations (Duffy, 2000). Avoiding the work load bias (2) was certainly achieved via raising the FAR rate (lowering the threshold for intervention) — easier access was obtained. Better decisions (1) were not made. The new performance indicators help illustrate the difference between these two objectives. (2) is reflected by where you sit on a ROC curve vs. (1) being reflected in the shape of the curve itself. Hopefully this kind of model will allow a more nuanced understanding of the dynamics around decision making to inform better responses to perceived issues.

5.3. The future for child protection

The three orders of issue facing child protection can all be mitigated if protection systems can be stabilised on an acceptable threshold for escalation and performance judged on all decisions made, not merely sentinel events.

The strategic goals become (1) stable decision making (not oscillating) and (2) improved overall performance (steeper ROC curve). Not the avoidance of all false negatives or false positives but rather, stable and equitable decisions that balance:

- Children at risk of maltreatment, who need risk detection services,
- Children already in our care who need support, and
- Families with other kinds of (non-statutory child protection) needs — who need not to be investigated.

Stable thresholds and judgment of true performance will help to mitigate poorly informed, sentinel event driven policy and calls for reform. Improved overall decision making may be possible, so uncertainty reduced and less errors of both kinds made. However, merely settling on an agreed intervention threshold in itself and judging the ability to achieve these targets will mitigate the second order and third order issues facing child protection agencies.

Policy and targets informed by the essential trade-off inherent in uncertain decision making will be less single issue focused and so there is likely to be less reform, and any reform that is contemplated will be less likely to be continually adjusting to proximate one-sided single issues (such as demand management or risk management) that lead to perpetually unravelling reforms.

New information on the value that intake decision makers already make can lead to better judgments of true performance and consequently reduce the pressure for continuous reforms. In the case of Child, Youth and Family, performance in responding to proximate issues has been remarkably responsive. When reformers wanted increased sensitivity to the risk of missing cases this was achieved. When the issue was demand management the number of investigations with no findings was reduced.

Protection agencies do not need to be 're-formed' nor do they need to try harder to perform better in the ways reformers have suggested. In New Zealand's case the agency has been demonstrably responsive to both the risk of missing cases and the risk of misallocating resources. It is the reformers themselves that need reform. They need to be better informed and better managed, not the business itself. There needs to be less judgment on the basis of sentinel events and greater emphasis on reframing the dialogue so that all parties can be part of a more intelligent and informed debate.

Debates about mandatory reporting verses a 'public health model' or 'inverting the pyramid' are both positions on a continuum within the same underlying system. We can have an emergency case level coercive response system and manage the level of investment if we can achieve an expected level of risk assurance (a specified intervention threshold). What the community is purchasing from child protection agencies a level of risk assurance, unfortunately with often unrealistic expectations about just how much risk assurance can be purchased. Child protection agencies need to have the knowledge capability to purchase an informed level of risk assurance, weighing the costs and benefits of the tertiary level responses with non-statutory responses.

That CYF's leadership in 2007 asked the authors what the FAR rate should be was to us our greatest measure of success. This was in the context of an agency that in 2004 would not admit that risk screening was part of the job or that errors were common or that false positives might be harmful (i.e. that there might be a limit to the number you were prepared to make).

5.4. Reframing child protection

Real change does not stem from restructures and poorly informed reactive responses to vaguely understood issues. Child protection systems need to be reframed, to receive better support and a more considered investment in them balanced against the need for the wider protective opportunities society has at its disposal.

By 'reframing' here we mean that better knowledge needs to be provided to support a better understanding and dialogue about the interconnected challenges that plague such agencies.

The primary aim of the decision performance model presented here should be to escalate the concerns and challenges facing front line social workers strategically in a way that improves dialogue between stakeholders. This should to some extent mitigate the risk of misunderstanding, poorly informed one-sided policy, and ill-informed criticism. These measures have already illustrated value in bridging understanding between front line social workers, Treasury officials and Ministers. In New Zealand the Treasury no longer criticises CYF for indulging in "soft demand";

> "We think [this] work is a leading example in the Public Service of how to construct high quality performance information. In particular, it makes clear the difficult trade-offs between managing demand and risk that are inherent in any child protection business. This clarity of information is a major step forward since the 2003 baseline review." (Letter from Secretary to Treasury, 18 December 2008, to Ministry of Social Development, p 1)

5.5. Adaptive leadership

This paper presents the technical component stabilising our child protection systems. It provides a decision outcome-based performance model to inform a change in the way we understand and manage our decision making service. It provides the ability to communicate the linked nature of our *seemingly* disparate strategic issues. However by itself it will not solve the challenge facing child protection agencies. By itself it will not mitigate the excesses of an overly zealous public, their ministers and media when high profile sentinel events occur.

As Heifetz (1994) suggests, the most difficult leadership challenges are not technical challenges. The real challenge is changing *the way we respond* to child deaths or demand crises when these occur. This is a different kind of leadership challenge. It is an 'adaptive leadership challenge' since it requires a change in understanding, culture and behaviour for all of our stakeholders.

> "The most common cause of failure in leadership is produced by treating adaptive challenges as if they were technical problems. … Adaptive challenges can only be addressed through changes in people's priorities, beliefs, habits and loyalties. Making progress

requires going beyond any authoritative expertise to mobilise discovery, shedding certain entrenched ways, tolerating losses" (Heifetz, Grashow, & Linsky, 2009, p 19)

We think the crisis of confidence in child protection has not been due to poor social work practice, a lack of technical capability or the inability to respond to emerging issues. What is required is leadership who can support the system to not over-respond to single events crises or poorly understood "demand" crises. Out leaders fail when they feel the need to be doing something such as build new processes in response to adaptive challenges. They will succeed when they build an effective system that still accepts that decisions sometimes fail. Successful leaders will directly confront the elephant in the room — that we cannot save every child.

The model provided here provides some support for capable adaptive leaders to stop the system tilting at windmills and instead engage in the difficult conversation. At the very least it will support such leaders to defend practice at the front line and take some of the pressure off social workers who would otherwise get blamed for not achieving the impossible when their job is hard enough as it is.

5.6. Re-orientating the response to concerns about maltreatment

The declared objective of this work was to enable better dialogue, understanding and management of the risk screening function in child protection. This is so that on the one hand child protection services do not spiral into wide spread and unfocused screening (and very little else), but also that we may still have an enabled tertiary level emergency response to young persons at immediate risk.

We have illustrated in this paper that there are departmental performance indicators that better reflect some of our core services and help to stabilise and make more sensible decisions about how to manage these – in this case the service was risk screening decision-making at intake to find child maltreatment.

We argued that child protection agencies require better design of strategically important processes to enable them to more accurately reflect and escalate the experience of front line social workers – in this case the uncertainty in intake decision-making – and so to better inform policy and service design.

Ultimately however, this new capability though necessary, is not sufficient. Measurement of our processes alone, no matter how central to our work, will not tell us how to target value. Our adaptive leaders also require ammunition of a different sort. We need the same kind of evidence based approach to communicating the needs of the children and families we see. Since 2008 we have been investigating the use of new kinds of indicators that illustrate the business of front line social work. By building a knowledge capability that tracks persons (and not services) and by text mining some of our 20 million case notes of the 400,000 New Zealanders who have been touched by our agency over 15 years we can start to show trends about kinds and levels of needs. This is beginning to show us the level to which we come across custody disputes. The level to which alcohol cuts across our work and cost of this on forward foot print of these families across child protection, health, income support and the Justice sector. By reorientating the dialogue towards the kinds of needs families are experiencing such as serious mental health issues, addiction and social isolation, this shifts attention back to what can be done to support these children and their families.

Rather than focus solely on questions about screening levels and abuse rates – which at the end of the day do not tell you what to do – there can be a dialogue about what health, corrections, police and NGO's are doing about the mental health, addiction and domestic violence in some of our families. Reformers, politicians, policy makers and thought leaders need to know what our social workers face everyday in a manner that makes sense to them. I.e. identifying trends and costs associated with different kinds of needs.

Social workers have an obligation to support the family in front of them. There is also an obligation to capture their stories in a manner that can provide strong voices for future generations in need.

We need to reframe child protection to have a stronger focus on what vulnerable children really need. We also need to understand better how tangible improvements in outcomes for them can be achieved. This will enable us re-orientate services and practice to make wise decisions about what can be done.

Acknowledgements

The authors thank Brenda Ratcliff for Heifetz, Louise Hull and Robbie Lane for editing, Paul Nixon for his sound advice on social work practice and tone. Thank you Ian Richards, Ray Smith, John Wardrop, and Peter Hughes for championing this work. Special thanks also to John Fluke, the ISPCAN Child Maltreatment Data Working Group and the American Humane Association for their advice, encouragement and support.

References

Australian Research Alliance for Children and Youth ARACY (2008). *Inverting the pyramid: Enhancing systems for protecting children.* Melbourne: Allen Consulting Group.

Benbenishty, R., & Chen, W. (2003). Decision making by the child protection team of a medical centre. *Health & Social Work*, *28*(4), 284–292.

Brown, M. (2000). Care and protection is about adult behaviour: The ministerial review of the Department of Child, Youth and Family Services. *Report to the Minister of Social Services and Employment, Hon Steve Maharey. Wellington*.

Child, Youth and Family (2000). *Child, youth and family submission to the government reviews of referrals and notifications and placement services.* Main Report. Wellington: Department of Child, Youth and Family Services.

Department of Health (UK) (1995). *Child Protection Messages from Research.* London: HMSO.

Department of Social Welfare (1999). *Annual report.* Wellington: Department of Social Welfare.

Duffy, A. (2000). *Child, youth and family submission to the government reviews of referrals and notifications and placement services.* Wellington: Department of Child, Youth and Family Services.

Fawcett, T. (2005). An introduction to ROC analysis. *Pattern Recognition Letters*, *27*, 861–874.

Gambrill, E. D. (2005). Decision making in child welfare: Errors and their context. *Children and Youth Services Review*, *27*(4), 347–352.

Gambrill, E., & Shlonsky, A. (2000). Risk assessment in context. *Children and Youth Services Review*, *22*(11–12), 813–837.

Heifetz, R. (1994). *Leadership without easy answers.* : Harvard University Press.

Heifetz, R., Grashow, A., & Linsky, M. (2009). *The practice of adaptive leadership.* Boston Massachusetts: Harvard Business Press.

Lonne, B., Parton, N., Thomson, J., & Harries, M. (2008). *Reforming child protection.* London: Routledge.

Mansell, J. (2006a). Stabilisation of the statutory child protection response: Managing to a specified level of risk assurance. *Social Policy Journal of New Zealand*, *28*, 77–93.

Mansell, J. (2006b). The underlying instability in statutory child protection: Understanding the system dynamics driving risk-assurance levels. *Social Policy Journal of New Zealand*, *28*, 97–132.

Melton, G., & Barry, F. D. (1994). *Protecting children from abuse and neglect: Foundations for a new national strategy.* New York: The Guilford Press.

Meyers, J. (Ed.). (1994). *The Backlash: Child Protection Under Fire*. London: Sage.

Ministry of Social Development, Child, Youth and Family and Treasury (2003). *Report of the department of child, youth and family services first principles baseline review, no. 5.* Wellington: Ministry of Social Development.

Munro, E. (1999a). Common errors of reasoning in child protection work. *Child Abuse & Neglect*, *23*(8), 745–758.

Munro, E. (1999b). Protecting children in an anxious society. *Health, Risk and Society*, *1*(1), 117–127.

Munro, E. (2004). A simpler way to understand the results of risk assessment instruments. *Child and youth services review*, *26*(9), 873–883.

Scott, D. (2006a). Sowing the seeds of innovation in child protection. *keynote address to the 10th Australasian Conference on Child Abuse and Neglect, Wellington, 14–16 February*.

Scott, D. (2006b). Towards a public health model of child protection in Australia, Communities. *Families and Children Australia*, *1*(1), 9–16.

Thorpe, D. (1994). Evaluating Child Protection Buckingham and Philadelphia. Open University Press.

Weeks, A. (1994). *A study of financial management practices in the children and young persons service in fiscal 1994.* Wellington: Department of Social Welfare.

Wood, J. (2008). *Report of the Special Commission of Inquiry into Child Protection Services in NSW State of NSW through the Special Commission of Inquiry into Child Protection Services in NSW*.