Australian Government
Australian Institute of Criminology

A tough nut to crack: Performance measurement in specialist policing

Zhivan Alach
Charl Crous

AIC Reports
Technical and Background Paper 53

A tough nut to crack: Performance measurement in specialist policing

Zhivan Alach
Charl Crous

AIC Reports
Technical and
Background Paper
53

www.aic.gov.au

ISSN 1836-2052
ISBN 978 1 922009 19 7

Published by the Australian Institute of Criminology
GPO Box 2944
Canberra ACT 2601
Tel: (02) 6260 9200
Fax: (02) 6260 9299
Email: front.desk@aic.gov.au
Website: http://www.aic.gov.au

Please note: minor revisions are occasionally made to publications after release. The online versions available on the AIC website will always include any revisions.

Disclaimer: This research report does not necessarily reflect the policy position of the Australian Government.

Edited and typeset by the Australian Institute of Criminology

A full list of publications in the AIC Reports series can be found on the Australian Institute of Criminology website at **http://www.aic.gov.au**

Foreword

Globally, police forces are increasing their use of specialist units within their organisational structure. Specialist units are generally either technical in nature (eg forensics), or operational (eg drug or fraud squads) and require innovative performance measurement frameworks to properly evaluate their effectiveness within the broader policing context.

In this report, the development of a performance measurement framework for Auckland Metropolitan Crime and Operational Support (AMCOS), a specialist policing unit of the New Zealand Police, is described. AMCOS encompasses a range of technical and niche units supporting policing operations in New Zealand. The performance framework, reflecting the roles and functions of the unit covers—forensic performance measures, operations support performance measures, intelligence performance measures and investigations performance measures.

The authors have written at a practical level that will assist practitioners to develop similar frameworks that can meet the needs of their specialist units, but that also reflects on analytical and theoretical aspects of performance measurement systems. The AIC's Technical and Background Paper series is aimed at methodological research that informs best practice and this paper will provide police services with an insight into the development of a successful framework that has assisted NZ policing to monitor and improve its performance.

Adam Tomison
Director

Contents

Tables

Acknowledgements

The authors would like to thank anonymous reviewers from the Australian Institute of Criminology for their insightful comments on earlier drafts of this report.

About the authors

Zhivan Alach was Policing Development Manager at Auckland Metropolitan Crime and Operation Support (AMCOS), New Zealand Police. He holds a PhD in Political Studies from the University of Auckland and has published widely on a range of military, security and policing topics.

Charl Crous is Superintendent State Intelligence Division, Western Australian Police. He holds a DPP from Charles Sturt University and has published articles on police intelligence and the management of Covert Human Intelligence Sources.

Executive summary

In the past two decades, public sector performance (and by extension police performance) has become increasingly important, especially in the Western world. However, while there is extensive academic work being done on the generalist tasks undertaken by most policemen every day, there has been very little attention paid to specialist policing performance. This is of some concern, as specialist policing presents a number of interesting challenges to the observer, particularly in terms of clearly identifying the role played by specialist policing in achieving larger outcomes. This report examines the issue of performance reporting in the specialist policing field and describes the development of an innovative performance framework for specialist policing by the Auckland Metropolitan Crime and Operational Support (AMCOS), a specialist policing unit of the New Zealand Police.

There has been increasing attention paid to public sector performance management in recent years, and many Western police forces have correspondingly become increasingly performance-focused, despite the challenges they face in doing so. The benefits that can accrue from performance measurement include improving value for money, improving managerial competency and increasing accountability (Collier 2006). But difficulties that relate to so-called 'perverse behaviours' can also arise (Loveday 2005: 98), where for a variety of reasons, performance measures become more important than the valuable activities they seek to describe.

Adopting performance measures for specialist policing has introduced its own set of challenges. Specialist policing is most easily defined by specifying what it is not—it is not the general, reactive patrol and investigative capability that comprises the majority of most police forces. Rather, specialist policing comprises two main categories—technical units, such as forensics and specialist operational units (eg helicopter, dog), and niche units, which are often investigative units engaged in proactive operations against a particular subcategory of criminality (eg drugs). The general lack of performance measurement attention that has been paid to specialist policing activities is likely related to issues of responsibility. Technical units provide a small part of a greater outcome and identifying what part they played in that outcome can be almost impossible. Niche units face even greater challenges, as their work can be lost among a much larger quantity of generalist activity. As such, it seems most rewarding to focus specialist policing performance measurement on outputs, at least until there is sufficient theory to take the next step to an outcome focus.

The AMCOS performance framework was developed locally, to guard against the imposition of a more generic model as well as promote flexibility and an emphasis on improvement rather than accountability. The framework, while focused on specialist policing outputs, is clearly linked to outcomes at the strategic level. The four main categories of the framework fall within the broader concept of technical and niche units described previously and comprise:

- forensic performance measures;
- operations support performance measures;
- intelligence performance measures; and
- investigations performance measures.

Identifying and developing measures for these categories was affected by different considerations, for example, the general absence of agreed definitions and reliable data on which to frame investigations performance measures and the contributory rather than absolute effect that forensics units have on major police outputs and outcomes.

Five factors were identified as being either critical in the development or implementation of such a framework. These were:

- the importance of managerial support;
- the value of consulting previous work,
- the necessity of consultation,
- issues around identifying respective shares of specialist resources to other policing groups, and
- recognition of ongoing difficulties with measuring the performance of niche units.

The next evolution of the AMCOS performance framework will focus on improving an understanding of the links between outputs and outcomes. AMCOS is currently working towards aligning and integrating performance measurement into its business planning, project management and risk management frameworks. The end goal is to have a centrally directed, but locally managed, performance management framework based on core strategic goals, which is integrated into planning and project management processes. This integration is operating on the principle that performance is only valuable if it serves as the basis for action.

Introduction

It has been said that professionalism in policing rests upon four pillars—accountability, legitimacy, innovation and national coherence (Stone & Travis 2011). In turn, the first three of those pillars rest upon knowledge of what police do, especially in terms of their performance. There cannot be accountability, legitimacy, or innovation until what has been done is truly understood. Numerous books and articles have been written on the topic of police performance, covering everything from complex statistical data, to the reactions of the public to police on the street (for useful introductory papers see Blumstein 1999; Braga & Moore 2003a; Collier 2006; Mackenzie & Hamilton-Smith 2011). British Home Secretaries have commented on the issue, noting that police performance 'is about delivering the best possible service to the public' (Home Office Police Standards Unit 2004: np) and given the attention that surrounds the annual publication of crime statistics, it is likely that the public also have strong views on how well the police are performing.

Amid the vast literature on police performance, however, very few works examine the issue of specialist policing performance—police intelligence groups, in terms of both intelligence collection and analysis; proactive investigation teams, such as drugs and organised crime; forensics units, including crime scene analysis; and advanced operational support, such as air, sea and canine capabilities. There are some studies of particular niches of the specialist policing environment (Canadian Police College Council of Investigative Excellence 2004; Mackenzie & Hamilton-Smith 2011; Willis, Anderson & Homel 2011), but they do not show links with other niches, or the difference between their particular topic and the general policing environment.

The purpose of this report is to show how AMCOS, a specialist policing unit of the New Zealand Police, has developed an innovative performance framework. This framework, which is continually updated, has followed the advice of experts and been derived from national strategic guidance, but shaped by local considerations.

In this report, the concept of specialist policing is defined and then some of the general difficulties surrounding police performance management are examined. The authors then explore the overall conceptual underpinning of the AMCOS system, before focusing more closely on the four main areas of AMCOS capability—forensics, specialist operations, intelligence and proactive investigations. Consideration is then given to describing both what has been done to date, as well as the improvements that will be made to the framework in the future. The report is deliberately focused at the practitioner level, but also includes mention of the various analytical and theoretical issues that have shaped development of the framework.

What is specialist policing?

Before one can talk about performance measurement for specialist policing, clarification is needed on what is meant by the term itself. It is a concept that can be difficult to be precise about and its particular characteristics will vary depending on the specifics of the law enforcement organisation in which such capabilities are based. Perhaps the easiest way to begin a definitional discussion is not to say what specialist policing is, but rather what it is not.

It is not the general public-triggered response/patrol capability that remains the core of most law enforcement agencies, especially in the Anglo-American tradition (Braga & Moore 2003a). It is also not the generalist, investigative capability that is usually the second largest component of law enforcement agencies. Both of these elements are reactive components whose level of demand is largely dependent on public requests and that usually have primary responsibility for the resolution of recorded incidents, whether they are serious crimes or disorder.

Specialist policing comprises two main types of units—those with very specific technical skills or capabilities who deal with a broad range of criminality, termed *technical units* (forensics, specialist operations and some intelligence units); and those with broader technical skills but who focus proactively, rather than in response to public requests, on a specific subset of criminality, termed *niche units* (proactive investigation units). Usually, a specialist policing unit will not have primary responsibility for crime control in a particular geographic area; it may be a sub-element of the group that has that responsibility, or it may well be part of an entirely different organisation. AMCOS fits the first type, while the FBI fits the second type. Often, specialist policing units will be multi-jurisdictional and will work across various law enforcement boundaries. They will often participate in multi-agency activities (Mackenzie & Hamilton-Smith 2011; Schneider & Hurst 2008) but will seldom be the lead agency.

From the above, it is obvious why technical units cannot be held responsible for crime in a particular geographic locale. A forensics unit cannot be held primarily responsible for crime resolution rates in a particular city, although its support is essential to such resolutions; and a dog unit cannot be held responsible for the number of car thefts although, again, its support is essential. They are thus contributory to, rather than responsible for, outcomes. With niche units, the issue of responsibility is more confused. Proactive units, such as drugs squads, will have primary responsibility for a particular crime type in a particular area, but there will often be a cross-over with generalist units, such as in the number of street-level drug arrests.

In the United Kingdom, the term *protective services* is sometimes used to refer to something quite similar to the concept of specialist policing (Flanagan 2008; O'Connor 2005) and in Australia, it often refers to police providing court security, prisoner transfer and at times, transport security officers. This report avoids the term for two reasons. First, it is counter-intuitive. For the average citizen, the term *police protective services* conveys the image of uniformed constables walking the beat on a Friday night, rather than specialist investigations into organised crime. Second, protective services includes homicide investigations and responses to disorder, which seem more akin to general response and investigation rather than fitting into the categories of technical and niche units.

AMCOS was designed to serve as a single repository for specialist policing services in the upper half of the North Island of New Zealand. Technical units include forensics components (forensic imaging, fingerprints and criminal profiling), specialist operational groups (dogs, air support and maritime units among others) and an intelligence collection and analysis section. Niche units include a number of proactive investigation teams primarily focused on drugs, organised crime and national security problems. Internationally, specialist policing units vary in detail (see NSW Police Force 2012; Queensland Police 2011) but the broad concept remains. As such, anything learned from measuring the performance of AMCOS has global applicability.

Parameters of performance measurement

Those with some knowledge of this topic may already have a question—if specialist policing units are not primarily responsible for crime control in a particular area, how can the popular measurements of recorded crime rates be used to evaluate them? This is indeed the key question, but before engaging with it directly, it is useful to set it in context by briefly examining issues of performance measurement in both the broader context of the public service and more specifically in relation to police forces. It is particularly important to understand the shortcomings and problems that surround the application of performance measurement to the police, as without such understanding, it is impossible to develop meaningful and useful performance frameworks.

Public sector performance measurement

Public sector performance measurement became an increasingly important issue in the Western world in the 1980s and 1990s (Carter, Klein & Day 1992; Fleming & Lafferty 2000; Schick 1996). Advocates of a new style of public service—sometimes termed New Public Management or NPM—hoped to propel what were seen as slow-moving, inefficient and overly bureaucratic organisations closer to a private-sector, corporate model, which would hopefully deliver better services for less money (van Sluis, Cachet & Ringeling 2008). Key to this push was accountability, which in turn required the development of performance measurement frameworks.

Performance measurement can be complex, so it is useful to clarify some core concepts. A standard performance classification scheme considers four elements—inputs, activities, outputs and outcomes. Inputs are the resources available to the organisation; the activities are the processes carried out by those resources. The outputs are the specific goods and services delivered, and the outcomes are the effect on the environment of those goods and services (Collier 2006). The total number of inputs used can be termed *economy*, the ratio between inputs and outputs *efficiency*, and the impact of outputs on outcomes can be termed *effectiveness* (Flynn 1986: 393). Before the 1980s, performance measurement in the public sector—although it was unlikely to be called that—largely focused on inputs, especially staying within allocated budgets. Over time, there was an increasing focus on efficiency indicators (Carter, Klein & Day 1992; Smith 1990). Today, it is generally felt that a focus on outputs and outcomes is of more benefit, especially when evaluating the quality of police work (Collier 2006). For the police in particular, the outcome of 'harm reduction' has become increasingly central (Mackenzie & Hamilton-Smith 2011).

Three broad benefits of performance measurement have been identified, namely improving value for money (efficiency), improving managerial competence and increasing accountability (Collier 2006). However, performance measurement can also have costs and not merely in terms of the resources required to undertake measurement. Smith (1995a) has identified eight negative effects ensuing from the publication of such data, including:

- tunnel vision—an emphasis on quantified elements of performance at the expense of other aspects;
- suboptimisation—the pursuit of narrow objectives at the expense of greater success;
- myopia—the pursuit of short-term success at the expense of long-term success;
- measure fixation—an emphasis on measures rather than underlying objectives;
- misrepresentation—the deliberate manipulation of data;
- gaming—the deliberate manipulation of performance to gain strategic advantage;
- ossification—an overly rigid system of performance measurement; and
- misinterpretation—misunderstanding performance data.

Flynn (1986: 389) has stated:

> At its worst, performance measurement has led to a concentration both on what is easily measured and what is susceptible to narrowly defined efficiency changes.

If targets are poorly defined and lack detail, problems reminiscent of those encountered in command economies can emerge (Smith 1990). Where there are too many indicators, however, there may be criticisms of unreliability, inflexibility and time wasting (Carter, Klein & Day 1992). Overly prescriptive indicators that specify not only what is to occur, but also how, can become divorced from the underlying objectives of the organisation. Thus, there is a need to carefully balance detail and prescription with freedom and flexibility.

While it is easy to assume that differing levels of performance by a given organisation (or sub-groups of that organisation) is primarily due to managerial competence, there are many reasons beyond the skills of a particular manager why those organisations (or sub-groups) might have differing levels of performance (Smith 1990). The organisations might have slightly different objectives, different needs, different costs, or might even measure performance differently. Separating out the impact of environmental factors (upon which any organisation has limited influence) on performance is a difficult task, but essential if the real value of management is to be identified.

The issue of 'window-dressing' is also important. Some writers suggest that performance measurement may sometimes be as much about the appearance of legitimacy as it is about instrumental improvement of performance (Collier 2008; Roy & Séguin 2000). This has been seen where organisations simply 'dress up' existing statistics as performance indicators (Carter, Klein & Day 1992), rather than going to the trouble of developing specific frameworks. At its most extreme, this may result in a performance version of creative accountancy, where two sets of performance measures are maintained—one for public consumption designed to ensure legitimacy is maintained and focusing on positive results, the other for internal use only (Flynn 1986).

The difficulty of measuring public sector performance can be illustrated by comparing the situation with that in the private sector. In the private sector, earning and profitability provide a convenient and simple 'bottom line' performance indicator. Despite that, private sector organisations have still developed large, complex sets of non-profit indicators (Bitichi et al. 2006; Carter, Klein & Day 1992), realising that such indicators present a far more accurate and nuanced picture of organisational performance. In the public sector, by comparison, there is no clear 'bottom line'—many important objectives are difficult to quantify and there is seldom an equivalent of earnings and profitability (Smith 1995a). Thus, if a good private sector performance framework—where there is a 'bottom line'—has to be complex, detailed and holistic—a good public sector performance framework has to be even more so.

Often, public sector activities are difficult to distinguish from one another, are produced in conjunction with other organisations and unfold over a lengthy period (Smith 1995a). Public sector performance is thus a particularly elusive concept

(Smith 1995a; Wisniewski & Olafsson 2004). This is especially the case with the police, where their goals are often complex objectives that cannot be achieved solely by police action (such as reductions in crime rates; Cockroft & Beattie 2009) and that are heavily dependent on the work of other agencies (Mackenzie & Hamilton-Smith 2011).

Police performance measurement

Around the world, but especially in the United Kingdom, there has been an increasing focus on police performance since approximately 1990 (Collier 2006; van Sluis, Cachet & Ringeling 2008). In that time, there has been a range of reports aiming to ensure that its approximately 43 police forces report on performance in similar fashion (Audit Commission for Local Authorities and the National Health Service in England and Wales 1998a, 1998b; Collier 2006, 1998; Home Office 2005, 2004, 2002; Public Services Productivity Panel 2000), with the goal being a centre-driven improvement in effectiveness (Home Office 2008; Home Office Police Standards Unit 2004; Loveday 2006, 2005). In 2007, performance measures were rationalised at the national level to focus almost entirely on public trust and confidence, and more recently there has been a further move towards devolving responsibility for performance measurement from the national to the force level (Barton & Barton 2011). The cascading of performance indicators from the national level to the police constable on the street has had mixed results (Butterfield, Edwards & Woodall 2004).

Statistical systems, such as COMPSTAT (first in New York and then further afield; Braga & Moore 2003b; Rosenbaum 2007; Schneider, Chapman & Schapiro 2009) have become increasingly common, focusing on the occurrence of specific crimes in limited areas over a particular timeframe (Stone & Travis 2011). In the Netherlands, a set of performance indicators for policing activities was introduced in the early 1990s (van Sluis, Cachet & Ringeling 2008). This was a major change for a police culture that had traditionally not been held particularly accountable for its actions or results (Hoogenboezem & Hoogenboezem 2005). In Australia, performance management was introduced into several state police services from the 1980s onwards (Fleming & Lafferty 2000), with Operational Performance Reviews (OPR) and similarly named reports deliberately emulating the COMPSTAT approach (Mazerolle, Rombouts & McBroom 2006).

As a result of this statistical focus, the rate of recorded and resolved crime has become the primary performance indicator for police around the world (Collier 2006; Dadds & Scheide 2000; Metropolitan Police Authority & Metropolitan Police Service 2009; New Zealand Police 2011; Western Australia Police 2011). In many police organisations, aggregated crime data are presented in league table formats showing the (perceived) comparative performance of different jurisdictions.

There has developed a debate about the applicability of simple, easy-to-use, numerical, 'New Public Management'-type performance schemes to the policing environment (Mackenzie & Hamilton-Smith 2011). Before discussing the shortcomings of such schemes, however, it must be remembered that this is a problem of police's own making—it was police forces around the world who embraced such simple, easy-to-use measures, preferring them to devoting the necessary resources to develop more rigorous, analytical and evidentially based frameworks. This is in direct contrast to military services around the world (Blumstein 1999), where there has been a substantial level of investment in Centres for Lessons Learned and operational analysis (United States Army 2011). Had police forces emulated their military counterparts and focused on studies of historical performance, they might well have developed the sort of doctrine and conceptual frameworks that have led to quantum leaps in effectiveness for some military services and that could also serve as the basis for better performance frameworks (Alach 2010a). As such, while some police forces do now appear to be increasingly focused on the quality of their performance measurement (notably in the United Kingdom where the constant evolution of measures has occurred), the lack of real investment in the field over the past two decades indicates that the situation is fragmented at best (Roy & Séguin 2000). There has been very limited investment in the sort of evaluation and research activities that are necessary to gain a

better understanding of police performance (Weisburd & Neyroud 2011) and little evidence of police forces truly implementing those learnings that have been gained from the few research and evaluation activities that have occurred (Bradley 2005; Chavez, Pendleton & Bueerman 2005; Lum 2009).

It is questionable whether standard police performance measurement schemes, with their overwhelming focus on crime rates (often at the expense of other aspects of policing activity) are relevant and accurate (Collier 2006). They can be particularly inaccurate when measuring the performance of police forces that are switching, or have switched, from a traditional, professional model to a community policing model (Braga & Moore 2003b). Further, while outcomes are usually regarded as central to performance management, commentators on police performance often emphasise that how police act is as important as what they achieve; police performance is as much normative as it is technical (Audit Commission for Local Authorities and the National Health Service in England and Wales 1996, 1993; Collier 2006; van Sluis, Cachet & Ringeling 2008). The legitimacy of policing is vital (Braga & Moore 2003a) and league-table type approaches cannot easily incorporate this aspect of performance; many actions such as mass random stop-and-searches that might improve performance vis a vis crime rates might actually harm police legitimacy. Related to this is the degree of alignment between what police perceive as good performance and what the public think; forces may focus on recovered and resolved crime rates, but the public may not perceive this as good performance (Kelling 1999), as they may continue to feel unsafe. Indeed, the public may well believe that any improvement in such statistics is merely the result of manipulation of recording practices by the police.

A narrow approach in performance measurement can lead to some of the negative effects cited by Smith earlier, particularly tunnel vision (Collier 2006). It can also lead to an over-emphasis on short-term targets at the expense of longer term objectives (Smith 1995a), despite the desirability of the latter (Collier 2006). The prioritisation decisions of police commanders will be influenced by the performance targets they are operating under, often leading them to devote the most resources to the most measured tasks, rather than those that may have more beneficial (albeit largely unmeasured) results (Davies 2000; Dupon 2003; Fleming & Lafferty 2000; Hoogenboezem & Hoogenboezem 2005; Loveday 1999; Vickers & Kouzmin 2001). A prescriptive approach to performance management can thus reduce the discretion of street-level police officers to best determine how to deal with a particular situation. This can, in turn, conflict with 'old style' police culture in which police discretion and flexibility is central (Butterfield, Edwards & Woodall 2004; Hoogenboezem & Hoogenboezem 2005).

The end result of performance measurement schemes can be 'perverse' behaviour, where performance targets or indicators become de-linked from the goals they are meant to achieve and instead become self-sustaining in their own right (Loveday 2005). Sometimes, the results are the opposite of those intended (Flynn 1986). This is more likely where the performance indicators focus on particular outputs, rather than outcomes or processes, which may 'offer perverse incentives to carry out those activities where it is easiest to notch up a big score' (Carter, Klein & Day 1992: 167). However, some seemingly perverse behaviour may not actually be such, as police outputs are often valuable in themselves (Braga & Moore 2003a), a point to which this paper will return later. Perverse behaviour can also be enabled by overly prescriptive performance indicators that do not allow for discretion or flexibility (Loveday 2005), as well as by performance indicators that are overly simplistic or mono-faceted and that fail to account for all relevant elements (Vollaard 2006).

Another key issue with police performance measurement is the differentiation between 'hard', ex-ante performance measurement, where targets are specified in advance and performance against those targets is measured strictly and softer, post-ante performance measurement, where performance over a period is evaluated in a more holistic fashion, incorporating more than just core performance indicators (Hoogenboezem & Hoogenboezem 2005). Some question the value of ex-ante performance targets (Vollaard 2006). Lawton (2005: 235) has stated that 'inspection, regulatory, and performance regimes that focus on prescriptive target-setting and the technical application of

pre-determined metrics ignore the importance of judgement'.

Perhaps because policing lacks the scientific foundation to set truly meaningful and feasible hard targets, most police performance measurement schemes around the world include an inspectorate function, where performance results are discussed and analysed. In New Zealand, this is executed by the Performance Group (Police National Headquarters) and in England, Wales and Northern Ireland by Her Majesty's Inspectorate of Constabulary. A problem with soft, post-ante performance measurement can develop when managers claim they are not responsible for poor performance, instead citing a range of causal environmental factors (Carter, Klein & Day 1992). While this may be partially accurate, it is unheard of for a manager to blame the environment for their good performance.

Another problem arising from prescriptive, ex-ante performance schemes is the potential for conflict in the applicability of performance targets at different levels of an organisation. For example, what may appear to be a priority (and thus a key performance target) at a national level may be largely irrelevant in a particular geographic locale (Loveday 2006), thus bringing into question whether performance in that field is a valid indicator in that place. At times, a performance indicator may be irrelevant not because it is a poor indicator, but rather because inadequate effort and analysis has been undertaken to convert that indicator into a meaningful measure at different hierarchical levels (and locations) of the organisation. Targets may also be deliberately unambitious—perhaps because it is realised that simple indicators are inadequate for real accountability and therefore it is best to set them at a level likely to be achieved—thus leading to suboptimal performance (Mackenzie & Hamilton-Smith 2011).

Related to the above dichotomy is the idea that performance measurement can include both accountability (whether 'hard' or 'soft') and learning elements. Accountability is past-focused and identifies whether what has occurred is good, bad, or in-between; learning instead focuses on how future performance can be improved by drawing on lessons from the past (Braga & Moore 2003b). Learning is not something that police forces have traditionally done well, except in the more limited field of technological advancement (Bradley 2005; Lum 2009; Weisburd & Neyroud 2011).

In an effort to overcome several of the problems cited above, Braga and Moore (2003a) have posited a comprehensive approach to police performance measurement. They state that 'controlling crime is the single most important core function of the police, (but) there are many other dimensions of performance that are valued' (Braga & Moore 2003a: 10). As such, they feel that any performance scheme needs to incorporate seven dimensions:

- reducing crime and criminal victimisation;
- calling offenders to account;
- reducing fear and enhancing personal security;
- ensuring civility in public spaces (ordered liberty);
- using force and authority fairly, efficiently and effectively;
- using financial resources fairly, efficiently and effectively; and
- quality services/customer satisfaction (Braga & Moore 2003a).

Braga and Moore (2003a) believe that it is important to measure performance in all of these dimensions. Ignoring one (or several) dimensions is at best a failure to fully appreciate the complexity of police work and at worst a contributor to the types of perverse behaviour noted earlier. This approach is therefore a particularly detailed, balanced scorecard. It takes into account the multiple influences acting upon police forces and the way in which performance in one area can involve trade-offs in another. In the last few years, there has been a limited degree of increased attention to such multidimensional approaches to police performance and the inclusion of additional factors beyond crime rates (Carmona & Gronlund 2003; Cockcroft & Beattie 2009; Hughes, McLaughlin & Muncie 2001).

Others have suggested that one solution to the performance measurement problems noted above might be to first identify what works—also known as best practice—and then measure the degree of adherence to that best practice (Lum 2009; Mackenzie & Hamilton-Smith 2011). This might be seen as a quality compliance approach. While this would be exceptionally useful when first setting

quality standards, as will be noted later, it is potentially troublesome as an overall solution due to its self-referential nature. It is also difficult to achieve, as Carter, Klein and Day (1992: 155) have noted, even in highly technical industries such as water management, where there is often clear scientific evidence (often lacking from the policing environment), 'standard-setting is the result of a political process that has to weigh up what is both desirable and what is feasible'.

A standards-based approach might quickly ossify (Smith 1995a) or instead become detached from the outcomes it seeks to achieve if environmental factors change. If best practice is too prescriptive, then flexibility and innovation may also be harmed. A similar approach of measuring milestones against a particular plan is also valuable in part, but again cannot solve all problems due to its self-referential nature and the likelihood of measure fixation developing.

Performance in specialist policing

When one moves from the general to the specialist—to the field of technical and niche units, as noted earlier—the difficulties in measuring police performance become even greater. First, outcome measures—even simple measures such as crime rates—are usually irrelevant for technical units and difficult to assign to niche units. Even more than with generalist policing, specialist policing groups will have large co-dependencies with other agencies or parts of the organisation (Mackenzie & Hamilton-Smith 2011); what is the 'outcome' of a fingerprint identification? Or, indeed, what is the outcome achieved by any technical unit? And what of counter-terrorism—if the measure is 'terrorist attacks' and the result is zero, how do we identify whether such was due to police actions rather than simple inactivity by terrorist groups? While the absence of activity can be reliably assumed to be at least partly related to police activity when there is a large enough sample size (such as crime rates; Vollaard 2006), when the sample size is a few incidents a year at most, the validity of assigning responsibility for any decrease (or increase) to police activities is more questionable.

The situation is often made even more difficult for niche units due to overlapping responsibilities. For example, AMCOS has primary responsibility for Level 2 and Level 3 organised crime in Auckland (as defined in the British National Intelligence Model; see National Centre for Policing Excellence 2005), but there are also District Organised Crime Units and a national Organised and Financial Crime Agency of New Zealand operating in the same space, albeit theoretically against different targets. If there were a reliable outcome indicator for the level of organised crime in Auckland, how could the respective effects of the different groups be calculated? The situation is the same in the United Kingdom and Australia, and especially so in the United States, where a range of metropolitan, state and federal agencies may all target the same range of organised criminal groups. Any simple outcome measurement of efforts against organised crime could therefore, at best, demonstrate the range of players involved and the overall effect of those actions; it could not, however, clearly identify the respective influence of those players. Carter, Klein and Day (1992: 32) have stated in relation to performance measurement that:

> ...the greater the complexity, the greater also is the scope for interdependence. The greater the interdependence, the more difficult it is to assign the ownership of performance to individual actors or agencies within the organisation.

In simple terms, technical units contribute to outcomes—but it is difficult to identify by how much. Niche units contribute to outcomes to a greater extent, but those results can often be lost amidst a much larger picture. Given these difficulties, it is the perspective of the authors that a meaningful specialist policing performance measurement framework should focus primarily (but not solely) on outputs and activities, with these two elements often blending into each other. This has the advantage of validity, as police have much more control over outputs than they do over outcomes (Dadds & Scheide 2000). Outputs and activities also have an inherent value in themselves (Braga & Moore 2003a, 2003b) and while there are problems with an output/activity-focused scheme, these can be mitigated to a certain respect, as later sections will show.

At the same time, where outcome measures can be validly assigned to specialist policing activities, they should be incorporated as part of a balanced

approach (Braga & Moore 2003a). However, this should only be done when the baseline of outputs and activities has been established; it is vital to develop the simpler elements of the framework before embarking on the more complex elements. This focus on outputs and activities is unavoidable given the current level of knowledge about the effect of police activities. It is therefore anticipated that in the future, given research initiatives like the Centre for AMCOS Lessons Learned, the framework can transition to one more focused on outcomes, but to do so now would be to put the conceptual cart before the horse.

There are further definitional conundrums to consider. For example, whether something is seen as an activity, output, or outcome will depend very much on who is doing the perceiving. As Blumstein and others have noted, police activities can be seen both as ends in themselves as well as contributors to other processes (Blumstein 1999; Braga & Moore 2003a). To use an earlier example, a fingerprint section will view the process of analysing a fingerprint as an activity (occurring within the section), the number of processes completed as an output (a service provided to something external to the section) and the successful identification and provision of that identification to an investigative unit leading to an arrest, as an outcome (altering the environment external to the section). From the perspective of the police as a whole, however, the arrest is an output at best and perhaps might be seen as an activity. For them, the outcome will be any changes in the crime rate related to the crime type for which that person was arrested. For niche units, outcomes will likely be measured in terms of prosecutions.

The second great challenge in an output/activity-focused framework is ensuring that the measures chosen are meaningful. If they are not meaningful, then the negative elements of performance measurement—particularly measure fixation and gaming—will swiftly emerge. Usually, an output-based approach attaches quantity and quality elements to each output category. Timeliness, sometimes seen as separate, is better seen as a facet of both quantity and quality aspects, given that any quantity is measured over time; the simple quantity of any output gives one indication of timeliness over a particular period. More specific elements, such as response within a particular period, can be incorporated into quality standards.

It can be relatively easy to specify meaningful outputs, such as 'the number of terminated drug operations per set standards'. Quality standards (perhaps focusing on best practice as noted earlier; Mackenzie & Hamilton-Smith 2011 then rest in a separate document, where they can be as detailed as required without making the performance framework unwieldy in itself. One potential approach to quality is to use a standards-based approach, similar to that used by the New Zealand Qualifications Authority; each particular level, that is 'Excellence' or 'A', has a specific list of defining characteristics so the evaluator measures actual performance against those lists in identifying the standard achieved. This might then lead to comments such as seven A-Grade terminations, four B-Grade terminations and three C-Grade terminations.

A more sophisticated aspect of meaningfulness is identifying whether or not particular outputs are responsible for the achievement of outcomes (Flynn 1986; Jackson 1993; Smith 1995b). There are at least three aspects to this:

- non-responsibility;
- differential success; and
- a lack of causality.

In short, non-responsibility involves a sub-component of a larger organisation delivering outputs that are necessary, but not sufficient, for the achievement of organisational outcomes, but which are intermediated through another sub-component before that outcome performance is achieved. The analogy of the 'widget factory' is useful. The factory produces the widgets (activity or output), but it is the sales staff that sell the widgets, gaining revenue (one outcome) and it is the overall structure of the company that determines profit (another outcome). One could easily have a situation where two companies produce widgets to the same standard and for the same cost, yet where one is profitable and the other is not due to differences in the quality of sales staff. Therefore, while widgets are partly causal to outcomes, they are not solely responsible. Holding the widget factory accountable for the overall performance of the company would be illogical (Flynn 1986). Similarly, in the delivery of particular

policing services, such as secondhand dealer checks, a lack of follow-up by other units can lead to a failure to achieve the desired outcome—in this case a decline in burglary rates.

Related to non-responsibility is the concept of differential success. This also involves the delivery of quality outputs coupled with a potential failure at the overall outcome level. The difference is that at least some outcome success is achieved, but only at a lower level; thus, there is differential performance. The Vietnam War is a prime example of this, where tactical victories (tactical outputs resulting in tactical outcomes) were not translated into strategic success (strategic outcomes) due to the absence of a coherent overall plan. This can be analogised to the police environment; terminated operations (tactical outputs) may affect drug availability in a particular locale (outcome success), but unless other groups also deliver quality services, then the overall outcome goal will not be achieved. One way of partly overcoming the problems of non-responsibility and differential success is to integrate activities, outputs and outcomes into a single plan, so that all of the contributors to overall performance are properly linked (Bratton 1999; Collier 2006; Mackenzie & Hamilton-Smith 2011; Smith 1995a). This approach is sometimes known as program logic (Duignan 2012).

The third issue, non-causality, is perhaps the most important of all and rests on limited knowledge of the link between outputs and outcomes. If a particular output X has no causal link with outcome Y, no matter how well we perform X, we will never achieve Y. It is these outputs that must be avoided at all costs—while the problems of non-responsibility and differential success can be overcome through better processes, non-causality can never be overcome. While there has been a substantial amount of research done on the causal link between police activities and the environment, there is no clear picture and no equivalent of the military's principles of war; we are still largely in the dark (Bradley 2005; Lum 2009; Stone & Travis 2011; Weisburd & Neyroud 2011). Police officers may assume that a particular activity, such as foot patrols, may lead to a particular outcome, but may lack the evidence to show that this is so. It may well be that such outputs are unproductive and merely take resources away from other, more beneficial activities.

In some situations, outputs that are causal to outcome success can be delivered poorly. In the case of the widget factory where there is a high fault rate, responsibility rests with the output provider. For the police, that would be the specialist policing agency. This might occur when a fingerprint unit has poor laboratory standards, leading to very few fingerprints being identified. However, in many other situations, outcome failure is not the responsibility of the output provider due to the issues of non-responsibility, differential success and non-causality noted above. Care must be taken to explore all of these issues before assigning blame for poor performance. Holding a manager accountable for an outcome when they do not control all of the elements contributing to that performance is illogical (Flynn 1986).

Overall, performance measurement for specialist policing is more difficult than for the police as a whole. Primarily, this is due to the fact that specialist policing provides specific services or outputs and is not primarily responsible for the achievement of policing outcomes. Any measurement framework must understand this and therefore focus on the output/activity level, while still remaining cognisant of the links between those outputs/activities and the outcomes they wish to achieve. It is also important to understand whether a seeming lack of correlation between the delivery of outputs and the achievement of outcome goals is actually due to the fact that those outputs are not causal or rather due to a shortcoming in overall strategy or structures.

The AMCOS performance management framework

The current AMCOS Performance Management Framework is approximately three years old (see Appendix for framework). It was initiated early in 2009 in response to a degree of criticism from within New Zealand Police as to the visibility of AMCOS activities. The rationale for its creation at the local level is succinctly expressed in the words of Sean Price, Chief Constable of Cleveland Police Force in England:

> Performance management is an area where there is no single accepted model for the police service, and any such model or framework needs to be fit for purpose (Home Office 2008: np).

The framework rests on a simple concept—first, identify what is important and second, work out how to measure what is important (Braga & Moore 2003a). Therefore, purpose determines the performance indicators chosen. The starting point for the framework was strategic guidance and specifically, the three Strategic Goals and associated outcomes listed in high-level New Zealand Police policy documents. From these, a number of outputs and activities were identified. Table 1 shows this first phase of classification.

The authors identified these contributory outputs and activities through an eclectic mix of literature

Table 1 Initial top level derivation of performance categories

Outputs	Activities
Evidence-based policing	Intelligence management
	Knowledge management
Timely and effective responses	Deploy to risk
	Manage organised and serious crime investigations
Thorough investigations	Forensic support
Strategic collaboration	Manage government relationships
	Manage agency relationships
	Liaise with other agencies

Source: APMF 2009

review, first-principles reasoning and consultation. This process of linking outputs to outcomes would not have passed scientific muster, but was the best possible at the time (given resource and time constraints) and as is noted below, has been substantially improved already. As such, the framework began as an output/activity framework linked to outcomes. This was done for the reasons already cited earlier in the paper on the difficulty of developing useful outcome measures for specialist policing activities without being overly prescriptive (see Vollaard 2006).

The value of such a strategic derivation seems obvious, but it appears to be relatively rare within police, although it is common in militaries. At least within New Zealand Police (and presumably overseas, given the literature surveyed), performance measures are largely presented in decontextualised fashion, separated from the strategic goals the organisation pursues except in the broadest possible terms, such as 'crime reduction'. There is seldom a clear hierarchical chain from outcome through enabler through output, activity and input. This separation makes it difficult to identify why a particular measure has been chosen and therefore divorces frontline practitioners from the strategic outcomes that should be their overall goal. The authors felt that an essential element in ensuring that any performance framework is understood and well-supported, as well as in ensuring that the framework remains outwardly focused rather than self-referential, was to make explicit the contribution to and links between low-level measures, and operational and strategic-level goals and objectives.

Former Commissioner of the New York Police Department, William Bratton, has talked about the value of a 'decentralized management system with strong strategic guidance at the top' (Bratton 1999: 16); the same applies to performance measurement. At the national level, the value is in identifying strategic performance areas, outcomes, impacts and Key Result Areas. However, this is not enough; lower levels must then analyse those high-level elements and distil from them the key measures relevant to their own business. The Home Office Police Standards Unit in England and Wales has said exactly this when it emphasised the importance of varying performance frameworks across—and by extension, within—forces to reflect differing circumstances (Home Office Police Standards Unit 2004). Alongside the top-down, strategic goal-driven process, there was a bottom-up element. This was based on an analysis of the literature, which suggested that performance schemes built from the bottom-up—based on the 'true nature' of police work—had value (Carter, Klein & Day 1992; van Sluis, Cachet & Ringeling 2008). Flynn (1986: 392) has stated that 'at their best, performance measures are developed by managers themselves to enable them to do their job better'. Stone and Travis (2011: 19), discussing professionalism in policing, note that 'careful analysis of local problems and the custom crafting of solutions continue to be necessary'; the development of a performance framework is indeed a solution to a particular set of problems relating to accountability and learning.

One quality of locally developed frameworks is that they are more likely to be perceived by those being measured as being focused on learning and improvement, rather than being strict accountability tools—'sticks' to use a colloquial term. Willis, Anderson and Homel (2011: 4) have noted:

> measurement systems designed to focus on performance improvements (as opposed to instruments of control) are much more easily accepted than systems designed exclusively for accountability purposes.

This is a particular problem given police traditional culture, which has often been seen as a challenge for New Public Management-style techniques and the accountability that accompanies those techniques (Butterfield, Edwards & Woodall 2004; Fleming & Lafferty 2000). By allowing staff to participate in the development of the framework, not only would the framework be better, but staff would also gain a sense of ownership and hopefully engage with it at a deeper level (Loveday 2006). Developing the scheme with input from the frontline also ensured that the scheme would remain focused on instrumental improvements to performance, rather than merely being a symbolic initiative designed to gain legitimacy for AMCOS (Roy & Séguin 2000). It was also anticipated that such an organic, bottom-up approach would help avoid neo-Taylorism—a separation between management (measurers) and workers (doers)—developing (Loveday 2006), which was felt to be likely if an

externally developed framework was imposed without consultation.

Another benefit of developing the framework locally was agility. Any changes could be quickly made at the AMCOS level, involving only 300 staff, rather than having to wait for movement at the national level, where there is a much larger bureaucracy. This would help make the framework particularly flexible and responsive.

The bottom-up element of developing the framework involved visiting all AMCOS units and discussing their work. From this, key activities and potential measures could be identified, and just as importantly, things that need not be measured could also be identified (Carter, Klein & Day 1992). These discussions were vital for two reasons:

- they helped ensure that any measures chosen were meaningful to the unit; and
- they also helped the unit understand why the framework was being developed.

As might be expected, there was some minor concern initially about 'counting everything', but AMCOS staff have increasingly become more engaged with the framework, seeing it as the best mechanism for expressing the degree of work that they do. Care was taken to ensure that the measures chosen were parsimonious.

The top-down and bottom-up elements were then blended together into an integrated classification scheme of AMCOS performance measures. Where the two elements met, there was a process of creative destruction and rationalisation. The core defining characteristic of each low-level element was identified and it was then placed against what was deemed to be the most fitting higher level category.

At times, the nature of the work done on the ground led to a slight re-definition of higher level categories, which being based on inductive reasoning sometimes failed to account for the evidence. If a higher level category seemed to have no available measures, it was left empty for the time being. An *Output Dictionary* was produced and disseminated, covering the scheme in its entirety. Table 2 shows a simplification of that output schema.

Over time, the Performance Framework has evolved (see Table 3). One key change has been a move away from Police Strategic Goals to Police Output Classes as the highest level of the framework (New Zealand Police 2009). This was done due to an increased focus at the national level on those outputs, and has had the benefit of forcing further analysis of the various measures and reconsideration of their best location within the framework. As a result, the Output Dictionary has been revised and updated several times. An example of the types of specific measures that exist under each broader activity category is shown in Table 4.

In developing the framework, thought was also given as to how to gauge the framework's success. Here, the thinking of Bititci et al. (2006) was crucial—is the performance scheme being used to regularly monitor performance and make decisions, and do staff see the system as valuable? Given the long-term nature of most AMCOS activities, it was intended that holistic formal reports would be produced every six months, with some performance indicators presented on a monthly basis. As noted further below, the framework is now transitioning to a quarterly tempo. It was also decided to survey staff on their opinion of the framework. The results of those surveys are covered later in the penultimate section on challenges and successes.

Table 2 Example of low-level measures identified and linked to higher level categories

Top level	Intelligence management		
Intermediate level	Basic products	Analytical products	Field intelligence processes
Specific measures	Notings	Strategic assessments	Gang events attended
	Operation-specific support	Tactical assessments	Other gang contacts
		Problem profiles	Community events attended
		Subject/offender profiles	
		Thematic assessments	

Source: APMF 2009

Table 3 Simplified representation of current derivation of activity categories from high-level guidance

Output classes	Specific APMF activity categories
Policy advice and ministerial servicing	Intelligence management
	Other knowledge management
General crime prevention services	Gather community intelligence
	Strategic collaboration
	Communicate knowledge outside police
	Code of conduct activities
Specific crime prevention services and maintenance of order	Specialist operational support
	Manage VIP and WitPro Ops
	Manage civil emergency responses
Police primary response management	Specialist operational support
Investigations	Forensic support
	Active investigation processes
	Terminated investigation processes
	Covert support
	Drug lab services
Case resolution and support to the judicial process	Criminal case resolution
Road safety program	n/a

Note: Framework has been modified to allow its replication here

Source: APMF 2011

Table 4 Simplified representation of measures linked to higher level activities

Drug lab services			
Deployment to suspected clan labs	Clan labs dismantled	Location assessments	High-value chemicals seized

Source: APMF 2011

Overall, by developing the framework internally within AMCOS, it was felt that several key benefits would result. First, it would ensure that the framework was accurate, meaningful and engaged with the unique nature of specialist policing. Second, it would ensure that staff would gain a sense of ownership over the scheme and thus it would be supported. Next, it would ensure that the scheme remained focused on real performance, rather than merely symbolism. Finally, it would allow the framework to be flexible and responsive to changing circumstances. As a result of all the above, it was felt that the framework would overcome many of the challenges noted earlier in relation to performance measurement in general, be less vulnerable to perverse behaviour and therefore be a better tool for management.

The performance framework in detail

The value of the AMCOS Performance Framework comes primarily from its overall structure, rather than its individual components. These individual measures are, for the most part, not unique. However, when linked with other measures and activities, one gains a better appreciation of the quality and quantity of specialist policing services delivered by AMCOS and their potential impact on policing outcomes.

In the following sections, the performance measures are described according to four specific areas that

fall within the broader concept of technical and niche units introduced earlier—forensics, operations support, investigations and intelligence. Before discussing those specific measures, however, it is useful to first consider some performance measures common to all AMCOS units, measures that are among some of the most innovative in the entire framework.

As was noted earlier, professionalism in policing rests upon a commitment to continuous learning and innovation, a willingness to reflect, and thus the creation and sharing of knowledge (Bradley 2005; Stone & Travis 2011). It is an area that police agencies have seldom been good at (Pendleton 2005). One common set of performance measures for all AMCOS units relates to the sharing of knowledge, both within and outside police, and the maintenance of good working relationships with governmental, non-governmental, and international partners and agencies. Knowledge sharing cannot and perhaps should not, be easily quantified; the number of meetings attended may be a simple indicator, but it lacks meaning. Rather, performance in these areas is reported on in a descriptive, qualitative manner, which provides a rich source of data for evaluation and comparison.

Another set of common measures relates to training and development activities, including levels of compliance with required training categories. It would be useful in future to expand this to something akin to a Military Essential Tasks List, which is a detailed method of identifying the level of preparedness and capability of a particular military unit (Global Security.org 2011; Tritten 1997).

Key to the learning and innovation aspect of the framework is the Centre for AMCOS Lessons Learned (CALL; Alach 2010a). CALL is an evaluation system that closely examines AMCOS activities, either individual operations or a closely linked series of occurrences. As a performance measurement tool, CALL is too resource intensive to be applied to all activities—a single investigative operation CALL report can take a month and AMCOS terminates more than 50 investigative operations a year—but it provides a useful adjunct to the broader, more easily reported elements of the framework. In particular, it contributes to our understanding of the relationship between outputs and outcomes (Boba 2003; Bradley 2005; Braga & Moore 2003a; Haberman & King 2011; Hughes, McLaughlin & Muncie 2001; Lum 2009; Vollaard 2006; Weisburd & Neyround 2011).

Because of this, CALL is vital to the future simplification and enhancement of the performance framework, not only for AMCOS and specialist policing, but also for New Zealand Police and law enforcement in general. This is because, as understanding of the linkages between outputs and outcomes improves, there is less worry about actual outcome measures. This may seem counterintuitive, so deserves further explanation.

Let us consider an outcome goal, Y. Measuring Y may be difficult, expensive and result in massive fluctuations over short time periods, thus making frequent measurement difficult. However, if we can be confident in a link between output X and outcome Y, then we can focus our measurements on output X, which is likely to be far easier to do, less expensive and more amenable to frequent reporting. It will still be necessary to occasionally confirm that the output–outcome relationship remains, but this can be done for auditing and reassurance reasons, rather than strictly for performance measurement. As such, the more CALL does to link specialist policing outputs with policing outcomes, the greater the future benefits for performance measurement are likely to be, both in effectiveness and efficiency.

Forensics performance measures

Forensic units fall squarely within the concept of *technical units* raised earlier, where the problem is that while contributory to major police outputs and outcomes, they are not responsible for such outcomes. As such, the challenge is to identify how much they have contributed and how important that contribution has been to overall results. The AMCOS forensics group consists of:

- forensic imaging (photography and videography);
- fingerprints; and
- criminal profiling.

AMCOS once included a DNA collection squad, but that has since been disestablished; however, performance measures for that squad were developed and are mentioned below.

Forensic imaging

For forensic imaging, performance measures focus on the number of jobs (outputs) attended where photographs are taken and/or video is collected and analysed, for both standard jobs (attended when resources are deployed per a standard roster) and emergency jobs, where an immediate response is required, such as a major motorway accident. It was felt that focusing on jobs was the best way of identifying the overall impact of the unit, rather than focusing on (for example) the number of pictures taken. The next phase for forensic imaging will be to shift towards some outcome (from the perspective of the unit) measures by examining courtroom and evidential results, such as the number of cases supported by photographic evidence and potentially even the number of cases where photographic evidence was vital to success. Gathering the data for these next-level measures will be very difficult, as cases are the responsibility of investigators separate to the forensic imaging unit.

Fingerprint unit

The fingerprint unit already had the basics of a performance measurement scheme in place, focusing on the life cycle of a print:

- the number of cases received (an indicator of workload);
- the number processed/finalised (an indicator of available resources);
- the number retained (an indicator of sample quality); and
- the number of identifications made, both via a computerised system and also via any priority suspect list provided by investigators.

The next step was to include the number of partial court case files prepared by fingerprint officers, who unlike forensic imagers are often required to testify themselves. Other additional measures included the number of crime scenes attended, as well as the actual number of physical exhibits examined. In the future, as with forensic imaging, new measures will be developed to identify the success rate of cases with fingerprint evidence against those without, as well as identify changes in efficiency (the number of cases processed correlated against staff numbers), as well as highlight areas of external concern (if the number of cases retained drops, it may indicate poor work at crime scenes by first attenders).

Criminal profiling unit

The criminal profiling unit was treated similarly to the forensic imaging unit, in that its primary performance measures were requests received (demand indicator) and products produced (primary outputs). It was decided not to split 'products produced' into separate categories, largely due to the fact that there is no formal typology of profiling products and as such, it is easier to merely list overall production and then describe the details separately. A second set of performance measures relate to the ViCLAS database system, the projected centrepiece of the unit into the future. These show inputs into the systems as well as outputs (potential linkage reports, both successful and unsuccessful). The next step will involve greater focus on courtroom results, which as might be noted, will be a common factor in future performance measurement across all forensics groups. It is likely that an in-depth CALL report will be required to develop the analytical basis for such indicators.

DNA collection squad

Performance measures for the now-disestablished DNA unit were simple—the number of samples taken (both voluntary and involuntary), and the number of offenders and crimes linked to those samples. Had the DNA unit remained part of AMCOS, future work would have focused on the evidential value of identifications.

Specialist operational support performance measures

Specialist operational services also fall within the *technical unit* category. From a thematic, if not organisational perspective, AMCOS has four full-time specialist operations groups:

- the VIP protection squad;
- the maritime unit;
- the dog section; and
- the air support unit.

It also has a number of part-time groups whose performance measures have not received the same degree of attention.

VIP protection squad

From a VIP protection perspective, the most obvious performance measure is a negative one, namely nil

harm to VIPs under protection. However, this is an unsatisfactory measure, because the absence of harm cannot be reliably construed as resulting from the presence of protection personnel unless the outcomes can be compared against the outcomes for a control group of VIPs who are not protected. Nor can the occurrence of harm automatically be assumed to be the result of poor VIP protection (ie non-responsibility). For example, someone shooting a VIP from a range of 1,500m with a military-grade sniper rifle is simply outside the ability of close-quarters VIP protection personnel to prevent.

Given this problem, outputs are the focus of VIP protection performance measures, namely the number of operations undertaken (this incorporates the fact that an operation is defined by particular quality standards, such as the publication of an operations order and adherence to standard operating procedures) and the number of person-hours of protection provided. There is no easy way to improve the measures in terms of outcomes. As such, any future enhancements of measures for VIP protection may borrow from Mackenzie and Hamilton-Smith (2011) and instead seek to 'benchmark' the unit against overseas best practice, which presumably has been derived from the evaluation of protection failures.

Maritime unit

The maritime unit presented great difficulties in identifying measures due to the sheer multiplicity of roles performed by the unit. It acts as a general duty response capability (in New Zealand Police terminology, a 'marine I-car'), a search-and-rescue (SAR) manager, an investigator into stolen outboards and similar equipment, and also works with other government agencies on lengthy, offshore patrols. As such, its performance measures are also varied. Inspiration from naval and air force performance schemes around the world led to the inclusion of sea hours as a measure. The number of jobs attended is another key measure, as is the number of people apprehended at or near the sea; this latter element therefore allows for the measuring of results from investigations into stolen maritime equipment. Further measures examine seizures of weapons and drugs, the first of which is particularly important given that foreign sailors usually have limited understanding of New Zealand weapon laws. The number of SAR operations managed, and people located and recovered is another performance measure. Lastly, the soft performance measures relating to liaison with other agencies are vital to the Maritime Unit and thus receive substantial qualitative description.

Dog section

Similarly to the fingerprint section, the dog section already had a performance system in place, focusing on deployments to jobs (divided into those in which a dog was used and those in which it was not), the number of apprehensions and arrests, the number of incidents cleared and the total number of offences cleared. While these measures are similar, they each have a subtly different nuance and therefore changes in their inter-relationships over time can be particularly informative. Another set of performance measures examine the number of times specialist detector dogs (drugs, firearms, explosives and bodies) have been deployed and is logically followed by measures of the quantity and type of drugs and weapons seized.

Air support unit

The air support unit (ASU) has a clear role, which made development of performance measures relatively simple. As with the dog section, the primary output is the number of deployments to jobs, followed by the number of flying hours provided. The next group of measures can be regarded as outcomes from the ASU perspective and include the number of offenders detected by both day and night, in both raw terms and also as a percentage of total deployments. This measure shows changes over time in the efficiency and effectiveness of the ASU and can stimulate more in-depth consideration of the causes behind any changes. The final performance measure for the unit concerns the number of times that the ASU has acted as a personnel carrier and deployed a unit such as the Special Tactics Group.

Next steps in developing specialist operational support performance measures

The next stage for both air support and dog performance measures will be a greater focus on outcomes and in particular, the added value of the two units. With a large enough sample size,

the difference in detection rates between live burglary (or other live criminal events, such as armed robbery) occurrences with and without air support can be identified; a similar calculation could be done for the dog section. While it is anecdotal, information received when the ASU briefly deployed to Christchurch in the aftermath of the recent earthquakes indicated that its influence may actually have been greater than is commonly perceived in Auckland; familiarity may well have bred a certain degree of contempt. Something akin to military-style Operational Analysis may be required to model the optimum mix of general duties branch staff, dog units and air support over a full range of policing scenarios. This might also include modelling the impact of advanced technologies, in that enhanced sensors and data links might provide a quantum leap in the degree of value provided by any sort of aerial platform.

For the other part-time operational units—SAR, specialist search, negotiation, emergency management and armed response—performance measures are focused on primary outputs, such as the number of operations undertaken and number of person-hours provided, as well as a small element of outcome focus in terms of drugs and weapons seized and the number of people located.

Intelligence performance measures

The final component of the technical unit category consists of intelligence units. The AMCOS intelligence performance framework is posited on Crous' (2010) two-fold typology of the internal activities of intelligence units (activities/outputs) and the effect of those activities (outcomes). The former, which primarily comprises intelligence products, is simple to measure. The New Zealand Police National Intelligence Centre has already set out definitions and quality standards for a range of analytical products and as such, it is easy to simply count the number of each type of report produced during a particular reporting period. Also measured quantitatively is the number of intelligence notings (a raw intelligence product) produced. Other output measures include the number of investigative operations that provided organic intelligence support, where an intelligence analyst brings their specific skillset into the investigation team.

The next step in measuring the internal aspect of intelligence performance will involve more closely tracking the eight fundamental elements often seen as central to good intelligence practice (Crous 2010):

- executive leadership;
- intelligence leadership;
- commitment;
- collaboration, coordination and partnerships;
- tasking and coordination;
- collection management;
- analytical capabilities; and
- training and education.

This is not performance measurement in the way described in the rest of the paper and so will not be described further here.

The second aspect of intelligence performance measurement focuses on outcomes and specifically, the difference that formal intelligence support has made in terms of outcomes for the police as a whole (Crous 2010). This is a difficult task and one that requires a substantial quantity of data that is not currently available; as such, this part of the performance scheme is as yet dormant. It is not enough to merely show that an intelligence-led operation has good results; this does not validly separate out the value of intelligence from that of other policing contributors. Ideally, intelligence-led and non-intelligence-led operations undertaken by units that are otherwise similar would be compared. Such comparative evaluations would help identify the relative utility of formal intelligence support and processes compared with informal, traditional approaches. There would also be a place for specific identification of those occurrences where intelligence made the crucial difference between success and failure.

Such a specific and focused approach to intelligence performance measurement is possible within AMCOS, as it would be easy in a relatively small group to either provide or not provide formal intelligence support to squads that are otherwise similar in terms of experience and expertise, and then evaluate the results. It would be difficult on a larger scale; in such cases it might be more useful to conduct a large scale evaluation of crime rates before and after the introduction of intelligence-led processes, corrected for other factors.

Once the impact that intelligence processes have had on operational outcomes has been identified, the next step is to identify the costs involved and thus identify the benefit-to-cost ratio of formal intelligence support. Anything less than this would give only a partial answer to the question of how formal intelligence contributes to policing outcomes.

Investigations performance measures

For the most part, developing a performance framework for forensics, specialist operational units and intelligence units—using the earlier typology, *technical units*—was relatively easy. Almost all had a good base of existing activity and output measures that could be better aligned with strategic guidance and combined into a coherent, holistic scheme and that led easily into potential outcome measures. However, much greater difficulties were encountered with the specialist investigations components focused on drugs, organised crime and national security targets (*niche units* using the earlier typology).

As noted earlier, measuring performance in this field is difficult due to the lack of agreed definitions and reliable data (Mackenzie & Hamilton-Smith 2011). Because of the relatively small number of such investigations compared with overall crime rates and investigations, statistical approaches lack some validity when measuring outcome effects; variations in operation numbers are much more likely to represent the level of police attention rather than the actual level of organised crime, for example. In drug law enforcement, existing mixtures of activity, output and outcome measures from varying perspectives (police and government) have led to a distinct lack of clarity (Willis, Anderson & Homel 2011) and limited progress towards the identification of any sort of best practice in terms of performance measurement.

It was decided that, as a start, the performance scheme for this area would focus on outputs and more specifically on operations. The term *operation* is commonly used to describe proactive investigations targeting a particular group of drug/organised criminals and thus had the value of common understanding. The next step was to divide operations into two types, borrowing from the British National Intelligence Model:

- Level 1 local operations; and
- Level 2/3 complex and organised investigations (HMIC 1997; National Centre for Policing Excellence 2005).

Level 1 operations are the vast majority of criminal investigations, involving small-scale criminality in a geographically focused area, whereas Level 2 involves cross-border offending and Level 3 involves serious, national and organised offending. Initially, a trifold division into Levels 1, 2, and 3 was attempted, but whereas distinguishing between Level 1 and 2 was relatively easy, distinguishing Level 2 from Level 3 was immensely difficult. As such, the core measurement for investigation teams is the number of Level 1 and Level 2/3 operations active as at the performance reporting date and the number of operations terminated (arrests made) during the preceding reporting period. The ratio between active and terminated operations in turn provides some information about the 'chunkiness' of investigation workloads. Linked to these primary outputs are measures relating to the number of arrests made, the quantity and type of drugs seized, and the quantity and type of weapons seized.

An outcome measure derived from the quantity and type of drugs seized is the amount of social harm to New Zealand avoided. This relies on an independently developed Drug Harm Index, which uses econometric analysis to arrive at a quantification of drug harm (BERL 2008). Another outcome measure used intermittently is the post-operation disruption assessment (Mackenzie & Hamilton-Smith 2011). These are assessments carried out once an operation has terminated that evaluate the impact that a particular operation had on the criminal environment, based on intelligence gathered during the termination and post-termination phase. Essentially the police equivalent of the military 'Battle Damage Assessment' (Diehl & Sloan 2005), these assessments are extremely valuable, but time consuming.

A next group of outputs in this area relate to judicial processes. They cover the number of trials completed during a reporting period, as well as the number of trials being actively prepared for. The number of witnesses managed during the period is also measured.

Future performance measures for investigative units will focus on two key elements—outcome harm reduction and investigative efficiency. In the first element, the primary focus will be better understanding the impact that police activities have on the availability of drugs (Willis, Anderson & Homel 2011). As such, substantial data on the price, purity and availability of various illicit drugs in the Auckland region will need to be gathered (Mackenzie & Hamilton-Smith 2011). A framework that has been developed in Australia (see Willis, Anderson & Homel 2011) appears to be an excellent base to build from, although in requiring a substantial amount of data from non-police agencies, it will require close coordination with information sharing efforts at the national level. If resources allow, more analysis into other effects of organised crime will also be undertaken, such as corruption and distortion of markets.

Investigative efficiency and quality is an area of performance that seems to have been largely ignored in the literature. This seems rather odd, as poor investigative techniques can have multiple negative effects, including wrongful convictions, wrongful arrests, avoidable acquittals and false prosecutions (Canadian Police College Council of Investigative Excellence 2004). These can, in turn, lead to Commissions of Inquiry and even the payment of substantial compensation to those who have been wronged (Canadian Police College Council of Investigative Excellence 2004). It therefore seems essential to factor in investigative efficiency, perhaps in the form of compliance with quality standards, into the future investigative performance framework (Braga & Moore 2003a). This may involve formally identifying some principles or key elements of investigative excellence (Canadian Police College Council of Investigative Excellence 2004) and then randomly auditing a number of cases at any one time to identify the degree to which actual practice matches best practice. There would thus be a requirement for continual research and analysis to ensure that any best practice criteria remain accurate and relevant.

Investigation support components also require performance measurement. Informant management is measured in a number of ways. At the first level of complexity are some basic activity measures, such as the number of informants recruited and registered. The primary outputs measured are intelligence releases. In the near future, an effort will be made to trace the impact that informant-derived intelligence has on investigations, a form of outcome measurement.

Other covert capabilities, such as technical and surveillance support, are also measured on the basis of outputs—Level 1 and Level 2/3 operations supported, and the total number of person-hours of support provided. As with informant management, the next stage of performance measurement for these units will be to identify the impact that they have had on investigative success through comparisons with operations in which such support was not provided.

A last set of measures covered under the investigative section could just as easily have been discussed under the section on intelligence. This set relates to intelligence gathering activities, also known as field intelligence, such as attending gang or community events. At this stage, these involve simple activity measures only and it is difficult to identify how these measures could be developed further.

Future development of the framework

There are several ways the AMCOS Performance Framework might be enhanced as a whole, above and beyond the individual enhancements noted above. Before describing these potential enhancements, it is important to briefly reiterate why a relatively small unit (300 out of 12,000 total staff in New Zealand Police) needs to develop and improve its own framework and how doing so counters many of the difficulties related to performance measurement.

As noted earlier, one of the values in developing performance frameworks internally is creating a sense of ownership (Loveday 2006). Another is meaningfulness. This is particularly the case for AMCOS, which is unique within New Zealand Police; while centrally developed measures focusing on crime rates and resolutions may be relevant across the 12 general purpose Police Districts, they are less so for AMCOS. Local development can also ensure

that managers involved in the production of the framework are more aware of the potential for those being measured to engage in perverse behaviour (Loveday 2005) and can intervene to prevent this occurring.

The next evolution of the AMCOS Performance Framework will be focused on enhancing cohesion, clarity and outcomes. Good performance measurement is facilitated by a coherent strategy (Collier 2006; Home Office Police Standards Unit 2004; Mackenzie & Hamilton-Smith 2011; Smith 1995a) and AMCOS is already working towards aligning and integrating performance measurement into its business planning, project management and risk management frameworks, using a program logic approach. The end goal is therefore a centrally directed, but locally managed, performance framework based on core strategic goals, which is integrated into planning and project management processes. Such a scheme is hoped to:

> provide the integrated management framework necessary to achieve the output and outcome performance required to fulfil organisational goals and objectives (Barrett 2000: 64).

Once fully implemented, the framework will enable true professionalism by facilitating learning from the past and thus improving the future. There will be enhanced clarity and accountability, as managers will only be evaluated against the performance indicators they have agreed to in the ex-ante planning process (Flynn 1986). It is also hoped to better link performance indicators with managerial incentives and controls, further clarifying expectations (Smith 1990). Internally, a truly integrated and meaningful performance measurement system may even lead to a more participative and consultative management style (Bititci et al. 2006), as managers continually seek information and insight into the reasons for changes in performance.

In identifying meaningful objectives for the planning process, the military-derived concept of Effects Based Operations (EBO) could be particularly useful (Alach 2010b). EBO is about identifying the effects required to achieve a particular goal, then tracing back the causal chains required to achieve those effects. As such, it rests on inductive reasoning backed up by research. By applying an EBO approach to any AMCOS strategy, the causal chains required to reach desired outcomes (strategic goals modified to the AMCOS geographic sphere) could be identified. These chains could then be turned into tiers of performance measures, some at the outcome/strategy level, the majority at the output/activity level and even some at the input level.

For example, one desired outcome for AMCOS might be reduced use of methamphetamine across Auckland. The effects required to achieve this might be:

- fewer first time users;
- a higher price per pure gram;
- reduced availability;
- a higher risk to consumers;
- improved social happiness; and
- better drug treatment.

From this, AMCOS-specific measures might be identified, such as increasing the price per pure gram by a certain percentage, seizing a certain amount of the drug and arresting a certain number of suppliers.

It is intended that the framework will be continually reviewed to ensure that measures do not become ossified. One mechanism of balancing the need for consistency (and thus comparability) over time with the need to be flexible has already been noted earlier—the use of quality standards. For example, the measure might remain 'number of operations terminated', but the standards used to define an operation might be modified to adapt to evolving circumstances.

Core to the future review of the framework will be rigorous analysis and comparative research undertaken by CALL. CALL will be able to build from post-operation assessments and identify elements of best practice. This will assist in the development of standards against which to evaluate future operations. CALL will also evaluate the outcomes achieved by particular activities and as such, help clarify the relationship between activities, outputs and outcomes, potentially simplifying and enhancing the performance measurement framework.

On developing the framework

The purpose of this section is to briefly discuss the experiences of the authors in developing the framework and identify some lessons that might be of value for others working in the field in either developing or critiquing similar frameworks. The theoretical challenges that abound within performance management have already been touched on; in this section, the focus will be more on the detail of implementing such a framework. Some key points should be noted:

- Management support. Any move towards innovation in performance management (which, by extension, may lead to criticism of units and staff) requires strong support from management. The authors were fortunate to have an innovative, future-oriented manager who was supportive of the new performance framework from the outset. Without support from the top, it is probable that staff would have been far less willing to provide the data required to construct the framework. When units delayed their reporting, communication from the Commander quickly cleared any roadblocks. The Commander was also supportive of the constant development and enhancement of the framework, rather than demanding that it stay in a fixed form. The support of the Commander was also vital to the success of the Centre for AMCOS Lessons Learned, a concept that seemed foreign (and perhaps a little frightening) to the staff being evaluated. Again, strong leadership and clear communication brought staff on board.
- A sound foundation. The Performance Framework was built on some existing work done between 2005 and 2007 to enhance AMCOS intelligence structures. As intelligence systems and processes were improved, there was a steady breakdown of information silos, which in turn, led to a greater focus on sharing knowledge and information. Without this initial work, the barriers encountered by the performance framework would have been much greater. The work on intelligence helped introduce staff to the concept of disruption, thus making the incorporation of this into the Centre for AMCOS Lessons Learned easier.
- Consultation. Too often, performance frameworks seem to be developed in an 'ivory tower', separate from the true nature of work (the authors grant that given their academic backgrounds, some readers may find the above comment to be unintentionally humorous). The authors countered this by visiting as many units as they could and talking through the issues with frontline staff. To ensure that staff feedback was continually incorporated, a survey of opinions about the framework was held at the end of 2009, after the first two performance reports had been completed. The framework was rated 3.4 out of 4 for usefulness and 3.1 out of 4 in terms of accuracy. Findings from this survey were then used to improve the next performance cycle. A

more recent survey (in 2011) asked for opinions on the latest version of the performance framework, now that it has been partially integrated into the planning process. This time, the framework was rated 9 out of 10 for accuracy, but only 6.8 out of 10 in terms of real value in measuring performance. This lower score is interesting. It is hypothesised that staff, having been exposed to a more advanced mechanism of measuring performance for two years, now have heightened expectations of the framework. It is hoped that future improvements to the framework noted earlier in this report will also improve the perceptions of staff and therefore increase survey scores. Surveying staff about performance (or indeed any sort of governance or management framework) is valuable, as the findings of such surveys are an excellent way to drive further improvements. If those developing performance frameworks do not discuss the framework with those being measured, they will likely end up developing a framework divorced from reality (and thus likely sub-optimal) and are almost certain to encounter staff opposition (and perhaps many of the perverse behaviours noted earlier).

- On a related point, at times, senior management may not be entirely sure what should be measured. Many managers have been brought up on a diet of crime rate data and may not have a clear idea what the objectives of specialist units truly are. This point reinforces the need for consultation at the grassroots, as frontline staff will be able to provide the necessary information for the development of meaningful measures.
- Proportional use of specialist resources. This point may be specific only to AMCOS, but is assumed by the authors to be universal. It relates to the share of specialist policing services provided to other policing groups. As police forces are heterogeneous, and traditional performance management schemes often stimulate competitiveness, there will often be keen interest as to the proportion of a particular service (fingerprints, dog deployments) provided to particular groups. Any framework must include some mechanism to identify these proportional shares and indeed, the authors would recommend that the more nuanced the data, the better. This allows for additional ad-hoc reports to be quickly developed in response to concerns or complaints. It also allows for the production of interesting analyses correlating crime rates in particular areas with the usage of particular specialist policing services.
- Theoretical issues of niche unit performance. The measurement of performance by niche units—proactive specialist investigation units in particular—is a problem that, to use the same analogy as this report's title is a 'nut that nobody has yet cracked'. It is the view of the authors that the only way to validly evaluate the effect of major investigative operations is to conduct operation-specific post-operational evaluations. These are incredibly resource intensive. The authors also believe that there is a need for substantial public surveying about the prevalence of organised crime. Also, understanding that the public may be reticent about commenting on such a topic, there should also be regular surveying of Covert Human Intelligence Sources as well. A combination of post-operational evaluations, public surveying and Covert Human Intelligence Sources surveying will enable us to begin developing real outcome measures for proactive specialist investigation units, in the organised crime field at least.

Those of an outcome-oriented personality who have read this report probably have a simple question—has the framework improved AMCOS performance? It is a difficult question to answer. Without going into sensitive detail, there have been substantial improvements in terms of the quantity and quality of outputs delivered by some units, although whether the visibility granted by the framework was the reason is unknown. There has been anecdotal evidence that increased reporting of operational workloads has stimulated enhanced esprit de corps. There is also evidence that some components of the framework, notably CALL, have had a directly beneficial effect on some tactics and procedures. The most telling piece of information comes from overall AMCOS Satisfaction Survey results in 2010 and 2011 (both carried out well after the Framework was first implemented). The AMCOS Satisfaction Survey asked recipients of specialist policing services how they perceived the delivery of those services. In 2010, the average score for satisfaction was 3.38 out of five; in 2011 it had improved to 3.82 out of five. It should also be remembered that any performance

framework, no matter how well-designed, is only as good as the use made of it.

Conclusion

In this paper, the development of the AMCOS Performance Framework has been described. AMCOS is a specialist policing body and specialist policing—whether forensics, specialist operations, specialist investigations, or intelligence analysis—is an area that traditional police performance measurement schemes do not deal well with. Indeed, some feel that traditional police performance measurement schemes do not deal well with any police work and instead, lead to perverse behaviours that have little or nothing to do with the true value of policing.

The AMCOS model was developed as a response to criticism of the visibility of AMCOS activities. It was derived from analysis of both strategic goals and tactical workstreams, and to date has been largely focused on outputs and activities, due to the fact that AMCOS contributes to but is seldom responsible for policing outcomes. Where AMCOS is responsible for such outcomes, however, some outcome measures have been developed. Performance indicators within the framework cover all aspects of AMCOS business and in the 36 or so months since it was first developed, the framework has been subjected to several major revisions.

The value of locally developed frameworks seems obvious—they are more easily accepted by staff, more difficult to manipulate and more accurate in terms of the measures they focus on. In the future, the goal is to increase the focus on outcomes within the AMCOS Performance Framework and combine it further with planning, project and risk management activities to provide AMCOS with the basis for true policing professionalism into the 21st century.

References

All URLs correct at July 2012

Alach Z 2010a. Building on the past: Establishing a centre for lessons learned within a police force. *Australasian policing* 2(2): 7–9

Alach Z 2010b. Policing and effects-based operations: Modelling methamphetamine. *Policing: An International Journal of Police Strategies and Management* 33(3): 490–505

APMF 2011. New Zealand police, AMCOS performance management framework. Unpublished document

APMF 2009. New Zealand police, AMCOS performance management framework. Unpublished document

Audit Commission for Local Authorities and the National Health Service in England and Wales 1998a. *Local authority performance indicators 1996/97: Police services*. London: Her Majesty's Stationery Office

Audit Commission for Local Authorities and the National Health Service in England and Wales 1998b. *The publication of information direction 1998: Performance indicators for the financial year 1999/2000*. London: Her Majesty's Stationery Office.

Audit Commission for Local Authorities and the National Health Service in England and Wales 1996. *Streetwise: Effective police patrol*. London: Her Majesty's Stationery Office

Audit Commission for Local Authorities and the National Health Service in England and Wales 1993. *Helping with enquiries: Tackling crime effectively*. London: Her Majesty's Stationery Office

Barrett P 2000. Balancing accountability and efficiency in a more competitive public sector environment. *Australian Journal of Public Administration* 59(3): 58–71

Barton L & Barton H 2011. Challenges, issues, and change: What's the future for UK policing in the twenty-first century? *International Journal of Public Sector Management* 24(2): 146–156

BERL 2008. *New Zealand Drug Harm Index.* Wellington: New Zealand Police

Bititci U, Mendibil K, Nudurupati S, Garengo P & Turner T 2006. Dynamics of performance measurement and organisational culture. *International Journal of Operations & Production Management* 26(12): 1325–1350

Blumstein A 1999. Measuring what matters in policing, in Langworthy RH (ed), *Measuring what matters: Proceedings from the policing research institute meetings*. Washington: National Institute of Justice: 5–10. https://www.ncjrs.gov/pdffiles1/nij/170610.pdf

Boba R 2003. *Problem analysis in policing*. Washington DC: Police Foundation

Bradley D 2005. *Tackling the knowledge deficit in policing: Strategic change versus ad hocery and paint jobs*. Paper to Safety, crime, and justice: From data to policy conference, Canberra, 6–7 June. http://www.aic.gov.au/criminal_justice_system/policing/~/media/conferences/2005-abs/bradley.ashx

Braga A & Moore M 2003a. *The 'bottom line' of policing: What citizens should value (and measure!) in police performance*. Washington DC: Police Executive Research Forum

Braga A & Moore M 2003b. Measuring and improving police performance: The lessons of Compstat and its progeny. *Policing: An International Journal of Police Strategies and Management* 26(3): 439–453

Bratton W 1999. Great expectations: How higher expectations for police departments can lead to a decrease in crime, in Langworthy RH (ed), *Measuring what matters: Proceedings from the policing research institute meetings*. Washington: National Institute of Justice: 11–26. https://www.ncjrs.gov/pdffiles1/nij/170610.pdf

Butterfield R, Edwards C & Woodall J 2004. The new public management and the UK police service. *Public Management Review* 6(3): 396–415

Canadian Police College Council of Investigative Excellence 2004. Achieving investigative excellence*.* Unpublished Discussion Paper

Carmona S & Gronlund A 2003. Measures vs. actions: The balanced scorecard in Swedish law enforcement. *International Journal of Operations and Production Management* 23(12): 1475–1496

Carter N, Klein R & Day P 1992. *How organisations measure success: The use of performance indicators in government*. London: Routledge

Chavez TD, Pendleton M & Bueerman J (eds) 2005. *Knowledge management in policing*. Washington DC: Office of Community Oriented Policing Services, US Department of Justice

Cockroft T & Beattie I 2009. Shifting cultures: Managerialism and the rise of 'performance'. *Policing: An International Journal of Police Strategies and Management* 32(3): 526–540

Collier P 2008. *Performativity, management, and governance*. Cambridge: Cambridge University Press

Collier P 2006. In search of purpose and priorities: Police performance indicators in England and Wales. *Public Money and Management* 26(3): 165–172

Collier P 1998. Operations and accountability: The role of performance indicators, financial devolution and strategy in the management of a police force. *International Journal of Police Science and Management* 1(1): 81–93

Crous C 2010. Conceptualising a performance framework for intelligence led policing. *Australasian Policing 2*(1): 35–41

Dadds V & Scheide T 2000. Police performance and activity measurement. *Trends & Issues in Crime and Criminal Justice* no. 180. Canberra: Australian Institute of Criminology. http://www.aic.gov.au/publications/current%20series/tandi/161-180/tandi180.aspx

Davies A 2000. Change in the UK police service: The costs and dilemmas of restructured managerial roles and identities. *Journal of Change Management* 1(1): 41–58

Diehl J & Sloan C 2005. Battle damage assessment: The ground truth. *Joint Force Quarterly* Issue 37. http://www.dtic.mil/doctrine/jel/jfq_pubs/1137.pdf

Duignan P 2012. *Program logic.org*. http://programlogic.org/

Dupon B 2003. The new face of police governance in Australia. *Journal of Australian Studies* 27(78): 15–24

England and Wales. Home Office 2008. *Guidance on statutory performance indicators for policing and community safety 2008/09*. London: Home Office

England and Wales. Home Office 2005. *Police performance assessments 2004/05*. London: Home Office

England and Wales. Home Office 2004. *Building communities, beating crime: A better police service for the 21st century*. London: Home Office

England and Wales. Home Office Police Standards Unit 2004. *Managing police performance: A practical guide to performance management*. London: Home Office

England and Wales. Home Office 2002. *Police Performance Assessment Framework*. London: Home Office

Flanagan R 2008. *The review of policing: Final report*. http://www.hmic.gov.uk/media/flanagan-review-of-policing-20080201.pdf

Fleming J & Lafferty G 2000. New management techniques and restructuring for accountability in Australian police organisations. *Policing: An International Journal of Police Strategies and Management* 23(2): 154–168

Flynn N 1986. Performance measurement in public sector services. *Policy and Politics* 14(3): 389–404

Global Security.org. 2011. *Chapter 2: Mission Essential Task List ((METL) development*. http://www.globalsecurity.org/military/library/policy/army/fm/25-100/chap2.htm

Haberman C & King W 2011. The role of research and planning units in law enforcement organizations. *Policing: An International Journal of Police Strategies and Management* 34(4): 687–698

Her Majesty's Inspectorate of Constabulary (HMIC) 1997. '*Policing with intelligence'. Criminal intelligence—A thematic inspection on good practice*. London: HMIC

Hoogenboezem J & Hoogenboezem D 2005. Coping with targets: Performance measurement in the Netherlands police. *International Journal of Productivity and Performance Management* 54(7): 568–578

Hughes G, McLaughlin E & Muncie J 2001. The permanent revolution: New Labour, New Public Management and the modernisation of criminal justice. *Criminal Justice* 1(3): 301–318

Jackson PM 1993. Public service performance evaluation: A strategic perspective. *Public Money and Management* 13(4): 9–14

Kelling G 1999. Measuring what matters: A new way of thinking about crime and public order, in Langworthy RH (ed), *Measuring what matters: Proceedings from the policing research institute meetings*. Washington: National Institute of Justice: 27–36. https://www.ncjrs.gov/pdffiles1/nij/170610.pdf

Lawton A 2005. Public service ethics in a changing world. *Futures* 37: 231–243

Loveday B 2006. Policing performance: The impact of performance measures and targets on police forces in England and Wales. *International Journal of Police Science and Management* 8(4): 282–293

Loveday B 2005. Performance management: Threat or opportunity? Current problems surrounding the application of performance management to public services in England and Wales. *The Police Journal* 78(2): 97–102

Loveday B 1999. The impact of performance culture in criminal justice agencies in England and Wales. *International Journal of the Sociology of Law* 27: 351–377

Lum C 2009. *Translating police research into practice*. Washington DC: Police Foundation

Mackenzie S & Hamilton-Smith N 2011. Measuring police impact on organised crime: Performance management and harm reduction. *Policing: An International Journal of Police Strategies and Management* 34(1): 7–30

Mazerolle L, Rombouts S & McBroom J 2006. The impact of Operational Performance Reviews on reported crime in Queensland. *Trends & Issues in Crime and Criminal Justice* no. 313. Canberra: Australian Institute of Criminology. http://www.aic.gov.au/publications/current%20series/tandi/301-320/tandi313.aspx

Metropolitan Police Authority & Metropolitan Police Service 2009. *Joint policing London: Annual report 2008/09*. London: MPA. http://policeauthority.org/Metropolitan/downloads/publications/annualrep2008-09.pdf

National Centre for Policing Excellence (on behalf of the Association of Chief Police Officers) 2005. *Guidance on the national intelligence model*. Bedford: National Centre for Policing Excellence

New South Wales Police Force 2012. *Specialist operations*. http://www.police.nsw.gov.au/about_us/structure/specialist_operations

New Zealand Police 2011. *New Zealand Police 2010–11 annual report*. Wellington: Minister of Police. http://www.police.govt.nz/sites/default/files/resources/annual/new-zealand-police-annual-report-11.pdf

New Zealand Police 2009. *Statement of intent 2009/10*. Wellington: Minister of Police

O'Connor D 2005. *Closing the gap: A review of the 'fitness for purpose' of the current structure of policing in England & Wales*. London: Her Majesty's Inspectorate of Constabulary. http://www.hmic.gov.uk/media/closing-the-gap-20050911.pdf

Pendleton M 2005. On the threshold of an innovation: Police and the management of knowledge, in Chavez TD, Pendleton M & Bueerman J (eds), *Knowledge management in policing*. Washington DC: Office of Community Oriented Policing Services, US Department of Justice: 1–32

Queensland Police 2011. *State crime operations command*. http://www.police.qld.gov.au/aboutUs/the_service/structure/scoc

Rosenbaum D 2007. Police innovation post 1980: Assessing effectiveness and equity concerns in the information technology era. *IPC Review* 1: 11–44. http://www.socialsciences.uottawa.ca/ipc/pdf/kr1-rosenbaum.pdf

Roy C & Séguin F 2000. The institutionalization of efficiency-based approaches for public service improvement. *Public Productivity & Management Review* 23(4): 449–468

Schick A 1996. *The spirit of reform: Managing the New Zealand state sector in a time of change*. Wellington: State Services Commission and Treasury

Schneider M, Chapman R & Schapiro A 2009. Towards the unification of policing innovations under community policing. *Policing: An International Journal of Police Strategies and Management* 32(4): 694–718

Schneider S & Hurst C 2008. Obstacles to an integrated, joint forces approach to organized crime enforcement: A Canadian case study. *Policing: An International Journal of Police Strategies and Management* 31(3): 359–379

Smith P 1995a. On the unintended consequences of publishing performance data in the public sector. *International Journal of Public Administration* 18(2): 277–310

Smith P 1995b. Performance indicators and outcomes in the public sector. *Public Money and Management* 15(4): 13–16

Smith P 1990. The use of performance indicators in the public sector. *Journal of the Royal Statistical Society Series A (Statistics in Society)* 153(1): 53–72

Stone C & Travis J 2011. Toward a new professionalism in policing. *New Perspectives in Policing* March. https://www.ncjrs.gov/pdffiles1/nij/232359.pdf

Tritten J 1997. Joint mission-essential tasks, joint vision 2010, core competencies, and global engagement. Short versus long view. *Airpower Journal* Fall: 1–39

United Kingdom. Public Services Productivity Panel 2000. *Improving police performance: A new approach to measuring police efficiency*. London: Her Majesty's Treasury

United States Army 2011. *United States Army Combined Arms Center—center for army lessons learned*. http://acac.army.mil/cac2/call/ll-course.asp

van Sluis A, Cachet L & Ringeling A 2008. Results-based agreements for the police in the Netherlands. *Policing: An International Journal of Police Strategies and Management* 31(3): 415–434

Vickers MH & Kouzmin A 2001. New managerialism and Australian police organisations: A cautionary research note. *International Journal of Public Sector Management* 14(1): 7–26

Vollaard B 2006. *Police effectiveness: Measurement and incentives*. Santa Monica: RAND

Weisburd D & Neyroud P 2011. *Police science: Toward a new paradigm*. Washington DC: National Institute of Justice

Western Australia Police 2011. *Annual report 2011*. Perth: Western Australia Police. http://www.police.wa.gov.au/LinkClick.aspx?fileticket=KEHjAU9lN5E%3d&tabid=935

Willis K, Anderson J & Homel P 2011. Measuring the effectiveness of drug law enforcement. *Trends & Issues in Crime and Criminal Justice* no. 406. Canberra: Australian Institute of Criminology. http://aic.gov.au/publications/current%20series/tandi/401-420/tandi406.aspx

Wisniewski M & Olafsson S 2004. Developing balanced scorecards in local authorities: A comparison of experience. *International Journal of Productivity and Performance Management* 53(7): 602–610

Appendix

AMCOS PERFORMANCE MANAGEMENT FRAMEWORK

INTRODUCTION AND OUTPUT DICTIONARY
V2 MAY 2011

AMCOS PERFORMANCE MANAGEMENT FRAMEWORK

- The AMCOS Performance Management Framework is a means by which staff can be made aware of the activities undertaken by AMCOS units.

- As it currently stands, it is **not** an evaluation tool; there are no targets or quotas.

- Also, as it currently stands, it does not measure **outcomes**. For example, it might measure the number of methamphetamine seizures, but it doesn't measure the effect of those seizures on the street price of the drug, the number of users, or the average purity of the drug.

- Much of what AMCOS does is not easily quantified. As such, the Performance Management Framework uses a lot of qualitative data.

- The AMCOS Performance Management Framework has recently been reviewed to better align with Police Output Classes.

HOW WE MEASURE PERFORMANCE

- The performance of AMCOS is measured against AMCOS Performance Outputs (APOs). These APOs are **output-type measures** derived from strategic guidance on **outcomes** and **impacts**. APOs exist at three levels: **operational measures**, **tactical measures**, and **sub-tactical measures**, depending on the complexity of the activity they describe The strategic guidance utilised to develop the APMOs included, but was not restricted to, the New Zealand Police Strategic Plan, New Zealand Police Annual Report, New Zealand Police Statement of Intent, and the Service Level Agreement with the Auckland Board of Management (BoM).

- We then visited units with initial drafts and consulted with those units to ensure that the outputs covered the range of activities that those units undertook. So far, there have been three major revisions of the output list, as we work towards

reducing the number of outputs, as well as ensuring accuracy.

HOW THE REPORTING CYCLE WORKS

- Currently, every six and twelve months, Policing Development produces a unit-specific information request covering activities undertaken during the reporting period. Some units report on more APOs than others, depending on the complexity of their work.

- Policing Development then collates the information returns and produces a performance report covering the reporting period.

RELATIONSHIP WITH BUSINESS PLAN AND UNIT PLANS

- Currently, we are working towards aligning the activities (outputs) described in the Business Plan with the Performance Management Framework. We are also working towards aligning the activities described in Unit Plans with the Framework. Eventually, it is hoped that all documentation will "sing from the same song book" and there will be a single, easily understood set of output classes that can be used for all major planning and reporting activities within AMCOS.

THE FUTURE

- AMCOS Policing Development has a five-year plan to enhance the Performance Management Framework. From mid-2011, we hope to include some **outcome** measurements, such as the price and purity of drugs. From mid-2012, we hope to know enough to include evaluation: that is, setting targets for each year and then seeing how well AMCOS does against those targets.

THE APOS

- **The following pages show the APOs in detail.**

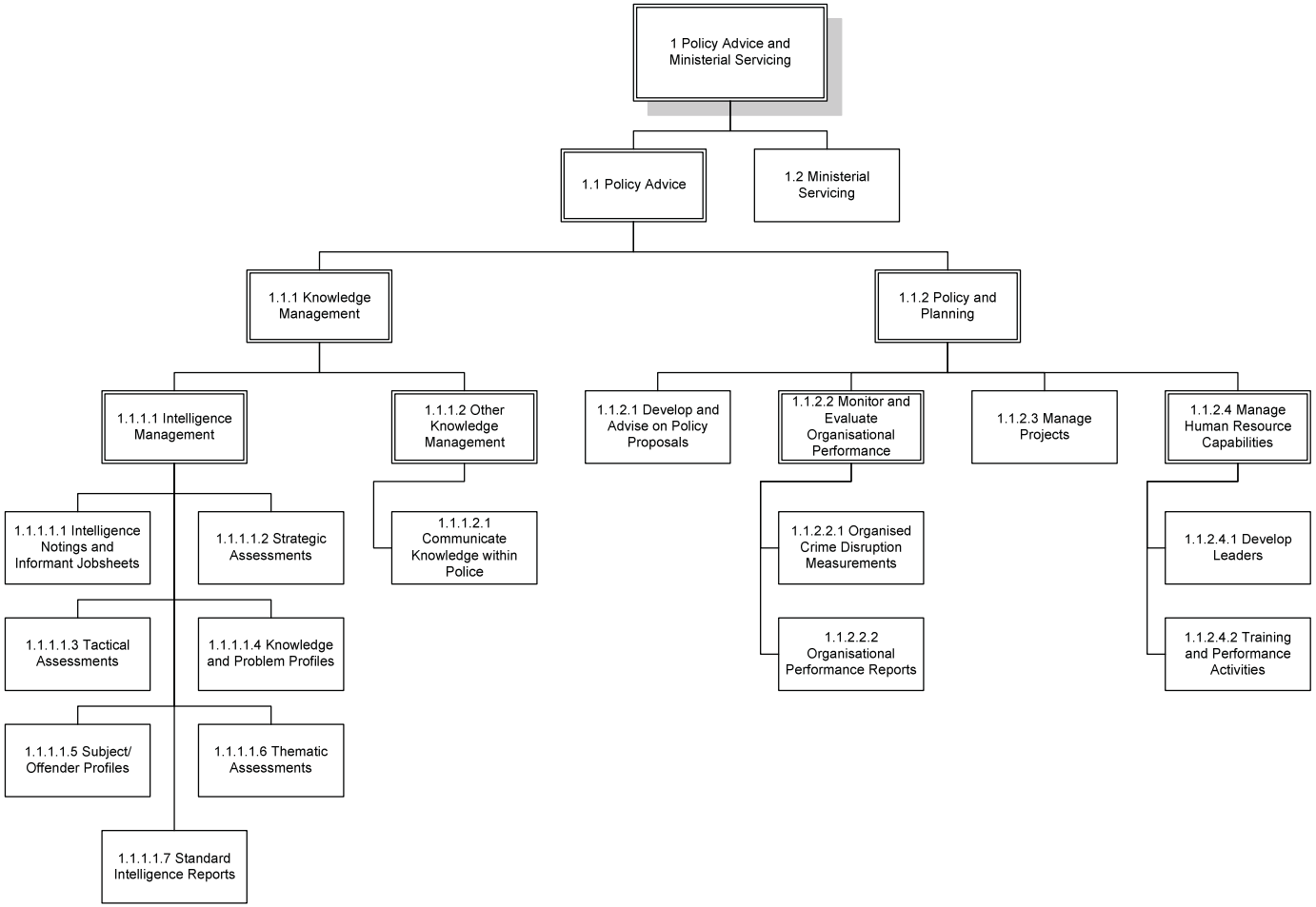

Figure 1 - APOs for Output Class One

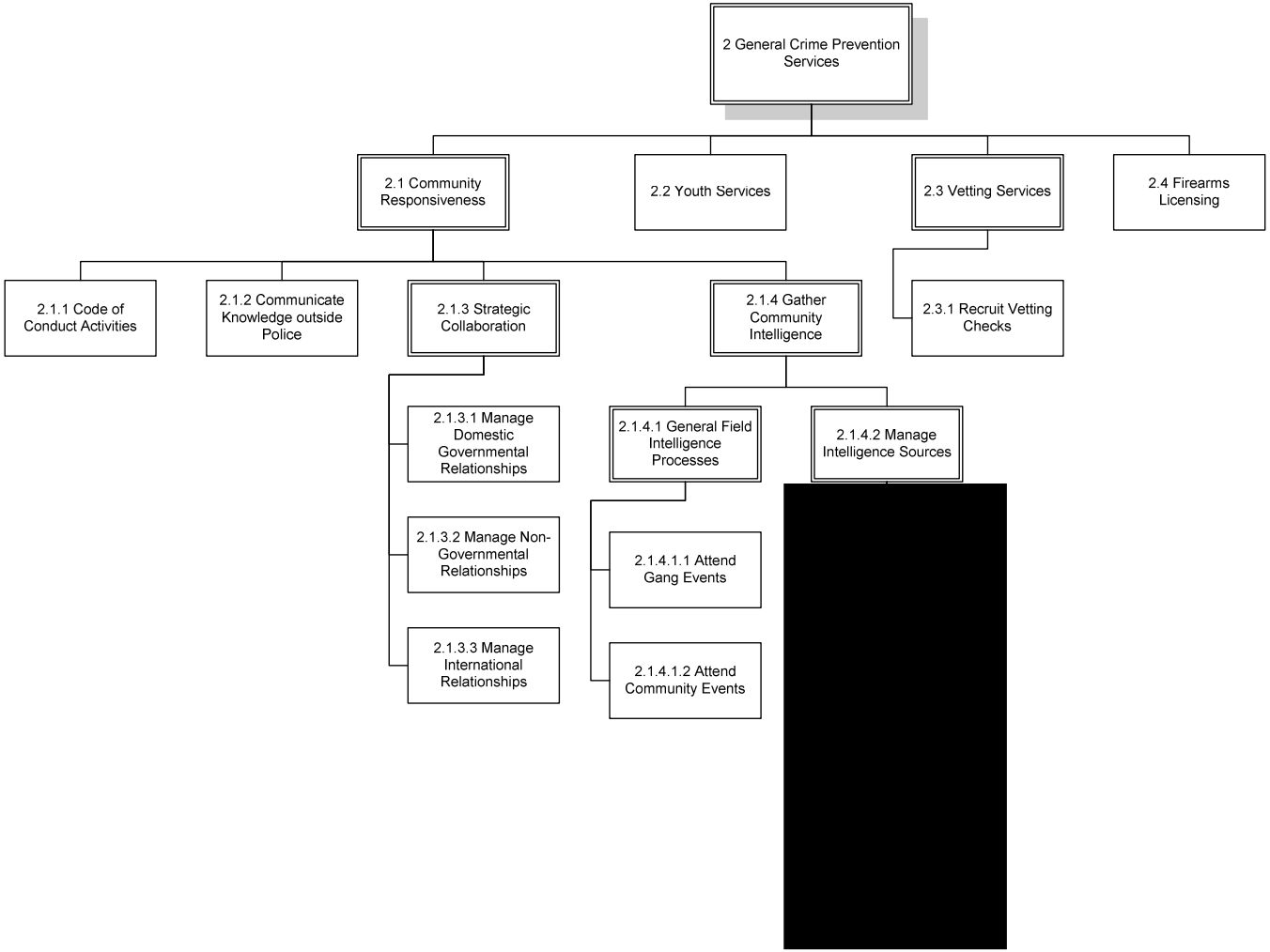

Figure 2 - APOs for Output Class Two

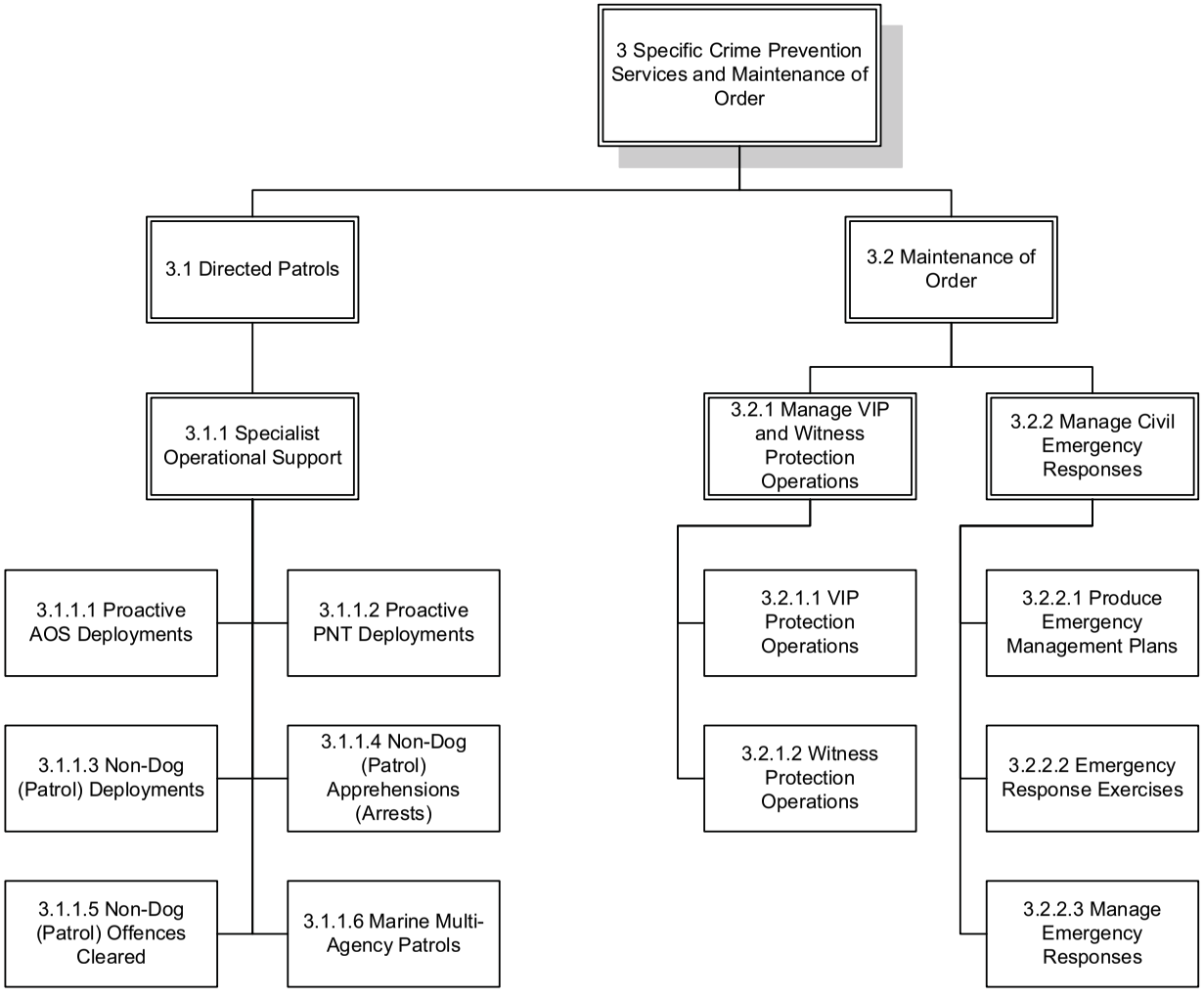

Figure 3 - APOs for Output Class Three

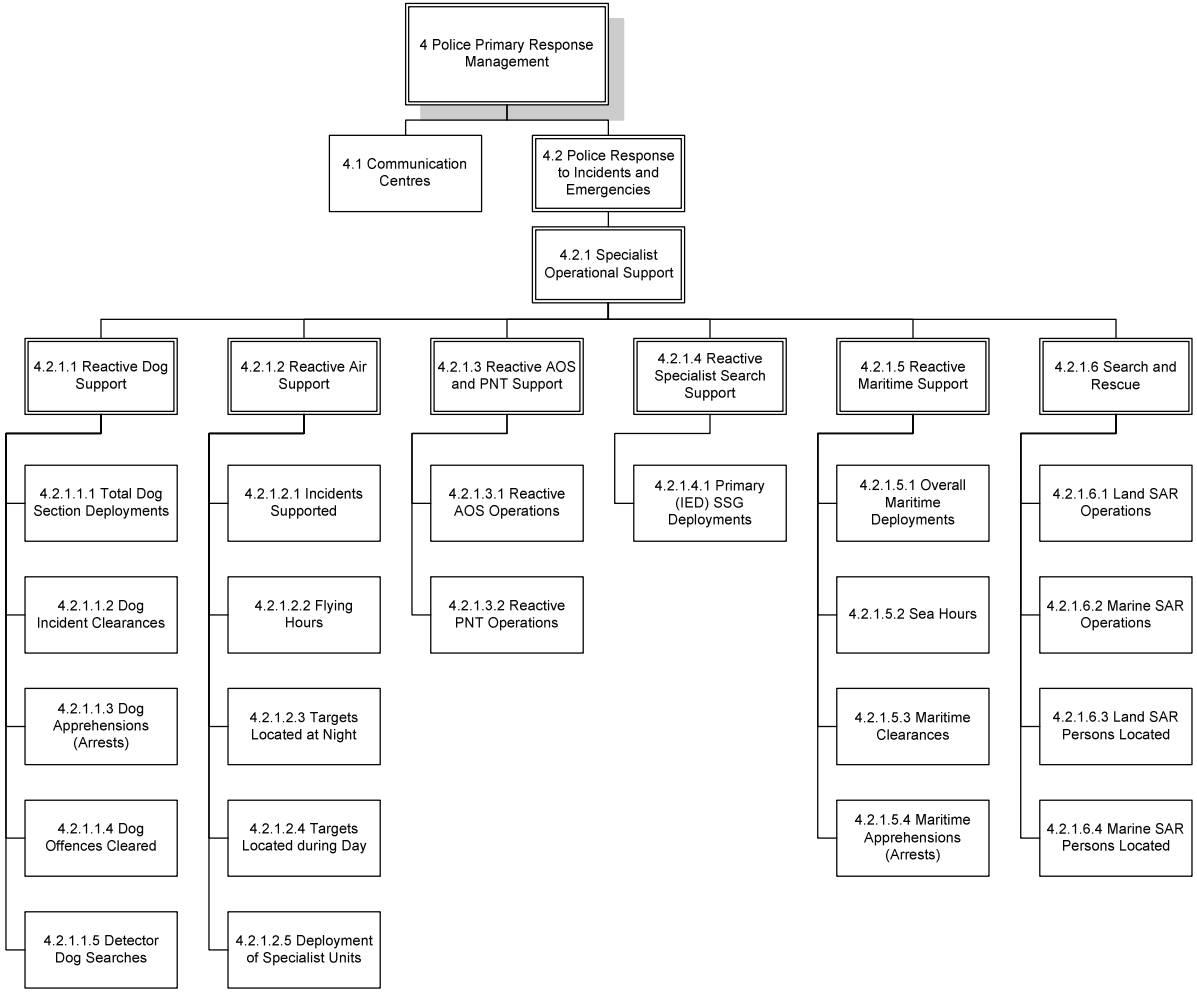

Figure 4 - APOs for Output Class Four

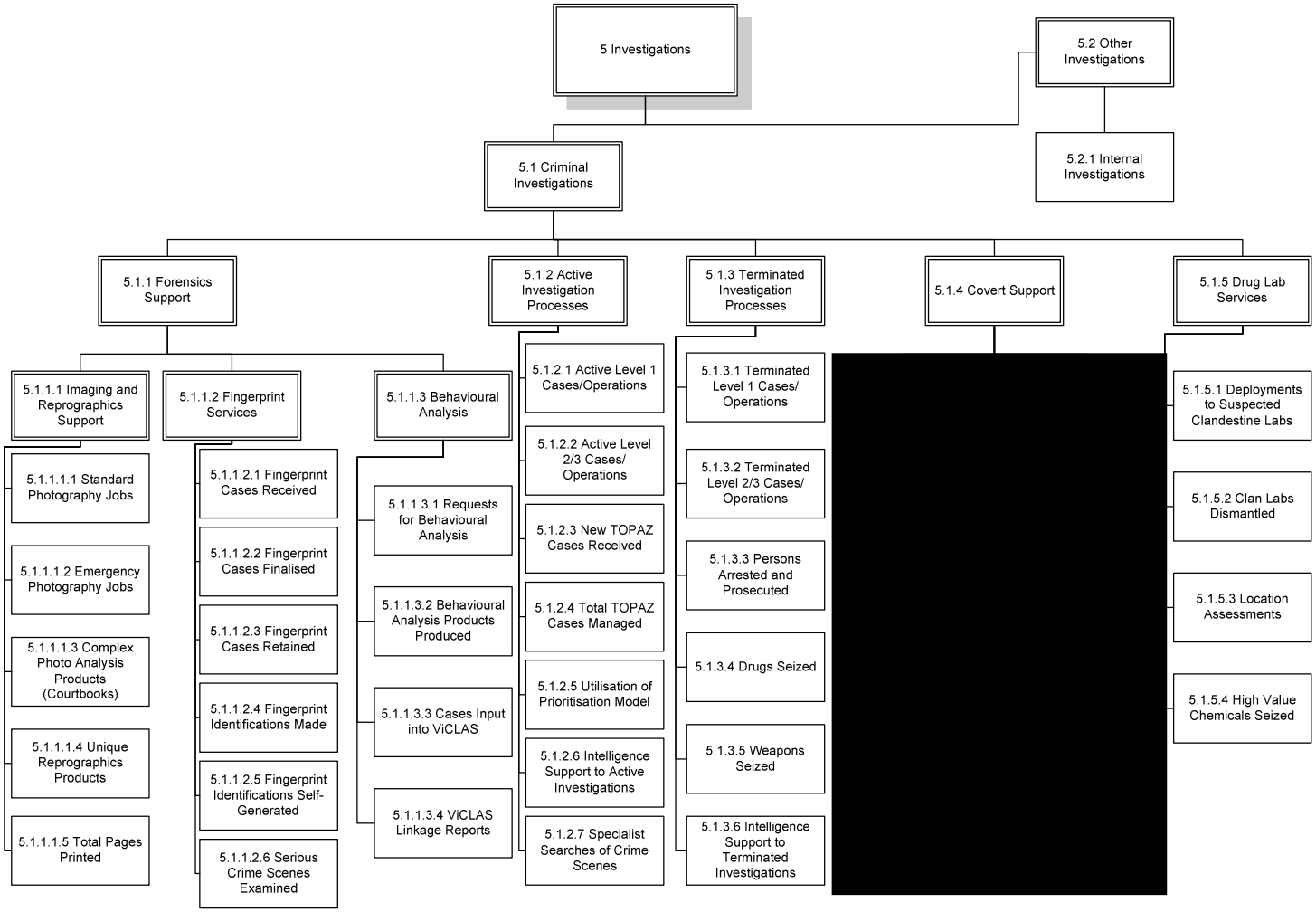

Figure 5 - APOs for Output Class Five

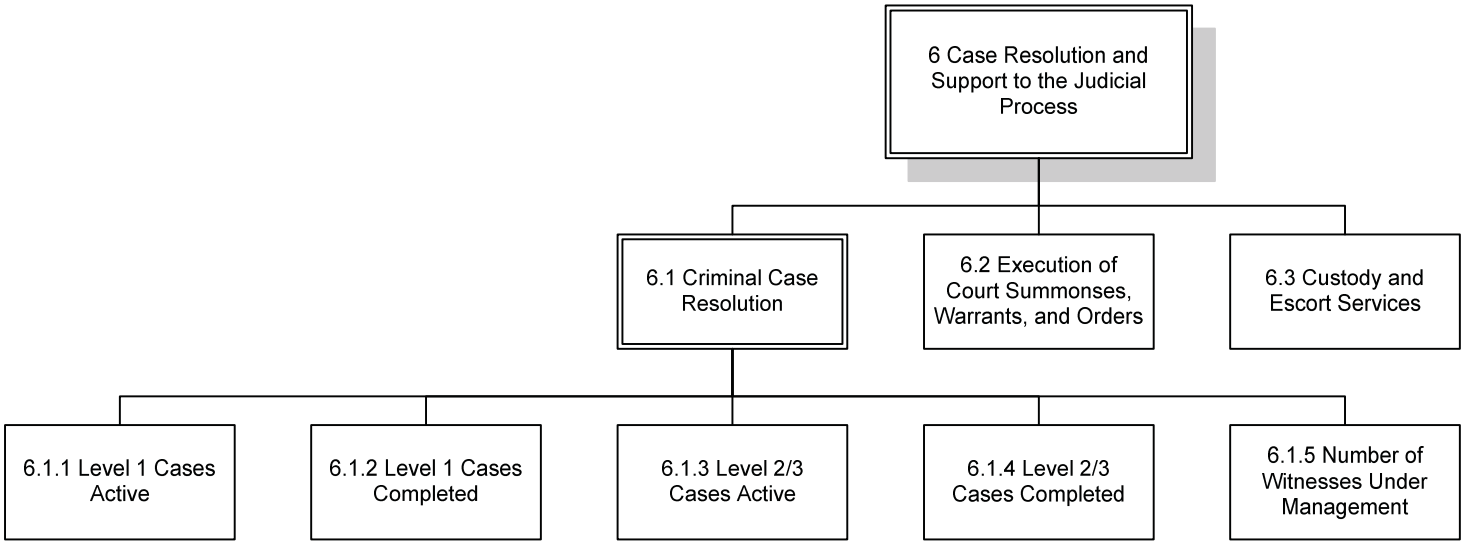

Figure 6 - APOs for Output Class Six