DAFF Biosecurity Quarantine Operations Risk Return
ACERA 1001 Study I
Performance Indicators
Report 1

Andrew Robinson, ACERA
Rob Cannon, DAFF Biosecurity
Robert Mudford, DAFF Biosecurity

acera
Australian Centre of Excellence for Risk Analysis

December 2, 2011

Acknowledgments

This report is a product of the Australian Centre of Excellence for Risk Analysis (ACERA). In preparing this report, the authors acknowledge the financial and other support provided by the Department of Agriculture, Fisheries and Forestry (DAFF), the University of Melbourne, Australian Mathematical Sciences Institute (AMSI) and Australian Research Centre for Urban Ecology (ARCUE).

The authors thank the following people who have provided invaluable advice and input: Victoria Allison (NZ MAF), Adam Beaumont (EPA Victoria), Professor Mark Burgman (ACERA), Michael Caporaletti (USDA APHIS), Matthew Chisholm (DAFF Biosecurity), Professor Neil Gunningham (Australian National University), Dr Fiona Haines (University of Melbourne), Steve Hutchison (ACCC), Tina Hutchison (DAFF Biosecurity), Stan Krpan (Sustainability Victoria), Kathleen Quan (DAFF Biosecurity), Bill Ross, Director Advanced Analytics, Australian Customs and Border Protection Service, as well as several other Customs personnel, Kari Sann (EPA Victoria), Robert Savage (DAFF Biosecurity), Professor Malcolm Sparrow (Harvard University), and Carolyn Whyte (ex-NZ MAF).

We are also grateful to two anonymous DAFF reviewers and a third anonymous scientific reviewer for very useful suggestions.

Contents

List of Tables

List of Figures

1

Executive summary

This report recommends a metric that can be used by AQIS (DAFF Biosecurity) as a indicator for its performance as an inspectorate. The Post-Intervention Compliance, PIC for short, is defined as the percentage of the units that arrive on the pathway that are compliant with quarantine regulations after quarantine intervention. We recommend that AQIS adopts PIC as an indicator of performance because

- PIC has a direct interpretation in the context of AQIS's mission,
- PIC comprises statistics that can be directly related to identifiable aspects of the inspectorate performance,
- PIC can be computed for many AQIS pathways using existing data holdings, and
- PIC is used by USDA APHIS and by NZ MAF as a measure of inspectorate performance for certain pathways.

We also outline a measurement strategy to collect the data that are required to compute the PIC.

2

Introduction

This report presents a recommendation for performance indicators for certain operations that fall under AQIS's purview. The performance indicators focus on AQIS operations as an inspectorate. We acknowledge that AQIS undertakes many operations to minimize biosecurity risk that are not related to quarantine inspections. For example, AQIS manages waste at airports, which involves the securing and disposal of human waste products from international aircraft. Another example of AQIS activity unrelated to quarantine inspection is the auditing of the production and supply chain of offshore importers. AQIS preiodically sends teams of auditors to examine the manufacturing processes of suppliers whose consignments are considered to be of high risk, for example because of the end-use. Some of those operations may be usefully monitored using the performance indicators that we advocate in this report although the decision to do so should be on a case-by-case basis.

AQIS describes itself and its operations on its website as follows:

> *AQIS manages quarantine controls at our borders to minimise the risk of exotic pests and diseases entering the country. AQIS also provides import and export inspection and certification to help retain Australia's highly favourable animal, plant and human health status and wide access to overseas export markets.*[1]

Performance indicators should reflect the goals of the organization that reports them. AQIS's goal is to minimize the risk of exotic pests and diseases entering the country. Ideally, the performance indicator should reflect the risk of exotics entering the country and also the effect that AQIS has upon that risk. This risk is unknown, as is the effect of AQIS's intervention upon the risk, but both can be estimated.

2.1 Definitions

In this report we will frequently refer to *pathways* over which AQIS has purview. When we refer to a pathway, we refer in the most general possible way to a sequence of actions that culminates in the entry of persons, animals, plants, or goods into Australia. We define pathways such that they can be nested within one another, or considered distinctly. For example, the entry of international air passengers is a pathway, as is the entry of goods by air cargo. The former pathway can be divided (*stratified*) into sub-pathways such as international passengers arriving in Melbourne Airport, or arriving on flight QF18, or all

---

[1] `http://www.aqis.gov.au`

passengers holding a passport from the European Union. We may refer to each of these examples as pathways, or as sub-pathways, depending on the context, mostly depending on whether primary interest focuses on the higher-level pathway or on the lower-level (sub-)pathways. In statistical terms, a pathway would be called a process.

We will call the fundamental components of these pathways *units*. This label encompasses international passengers, consignments of cargo, and so on. For the cargo pathway, we will refer to a consignment as being the unit, rather than the individual commodities that comprise the consignment. Hence the inspection of 600 pieces of fresh fruit from a consignment would be counted as the inspection of a singles unit.

We will also refer to the *inspection* of units or pathways. An inspection of a unit will be the examination of the unit for non-compliance. The inspection may be a multiple-step procedure, for example it may include screening of units using information such as accompanying paperwork, or the inspection history of similar units. The performance indicator that we will later discuss provides a way of assessing the utility of the screening as part of the inspection process, if the inspectorate wishes to consider the screening process separately.

Non-compliance is the general label for the state in which a unit is not compliant with quarantine regulations. The types of non-compliance that the inspection is intended to identify will depend on the pathway. Examples include: biosecurity risk material (BRM) in cargo consignments, *undeclared* biosecurity risk material (UBRM) carried by international passengers, and inadequate documentation for cargo or arriving vessels. Non-compliance can take a number of different forms, for example, BRM in cargo consignments could refer to a consignment of BRM, a consignment of non-BRM units contaminated with BRM, or a consignment of non-BRM units where BRM items have been hidden inside. Each of these would be treated differently and the risks associated with each are different.

The 'inspection of a pathway' will be used as a short-hand phrase for the inspection of units on the pathway. The inspection may be of only a *sample* of the units on the pathway, or it may be of all the units on the pathway. The *inspection count* will be the number of units on the pathway that are inspected, and the *volume* of the pathway will be the number of units on the pathway.

Given the inspection outcomes, the *approach count* will be the number of non-compliant units on the pathway, and the *approach rate* will be the proportion of units that are not compliant. The *leakage count* will be the number of units that would or should have been intercepted by intervention but were not, and the *leakage rate* will be the proportion of units on the pathway so defined. Finally, the *inspection effectiveness* will be the probability that a non-compliant unit that is inspected will be intercepted.

2.2 Performance Indicators

A performance indicator should provide a single assessment of the overall situation, but should also be decomposable into distinct and concrete elements of the regulator's operation. An indicator can be used to predict the effect upon the regulator's performance of different kinds of investment. In AQIS's case, it might be useful to be able to compare the predicted effects of increasing the number of inspections against public relations exercises that may reduce the rate of non-compliance.

AQIS has a number of substantial advantages with regards to measuring its performance, compared with some other regulators, for example, the Australian Competition and Consumer Commission (ACCC) or the Environmental Protection Agency of Victoria (EPA Victoria). We now examine the two most substantial advantages.

First, AQIS can obtain information about all of the activities of interest within many of the pathways that it manages. For example, a list of all international vessels arriving in Australian ports is available to AQIS. In contrast, the ACCC is responsible for ensuring the safety of a number of product lines, for example, small plastic novelty toys. Obtaining a list of all the producers and suppliers of small plastic novelty toys is an extremely difficult task. Similarly, the EPA is responsible for monitoring not only the environmental performance of some 700 licensed premises but also the aggregated effect of all unlicensed premises. Again, monitoring the performance of organizations that are only indirectly linked to the EPA is a challenging task.

That is not to say that AQIS has full information for all the pathways that it is responsible for. For example, information about commercially imported and reported goods arriving by shipping container that may be eligible for quarantine oversight comes to AQIS via ICS, which is the computer system operated by the Australian Customs and Border Protection Service (Customs). Only those entries that are flagged by an appropriate profile are made available to AQIS. Therefore the inspectorate relies on both the accuracy of the profiling system and the accuracy of the recorded information to determine which activities are in scope. If either of these elements is questionable then activities over which the inspectorate has purview may be unmonitored.

Second, the ACCC and EPA recognize a great variety of types of non-compliance. Whilst this recognition is important for developing a nuanced position towards the processes being regulated, it also greatly complicates the measurement and analysis of performance data. AQIS as an inspectorate presently classifies activities as “compliant” or “non-compliant”, although it recognizes that there are degrees of non-compliance, for example seizures from international passengers are classified as being *risk* or *high risk*. Such a simplification of the operating environment facilitates producing and reporting meaningful performance indicators. The disadvantage of the simplification is that it complicates the prediction of consequences, because a wide range of consequences can result from non-compliance.

2.3 Background

We begin with a brief history of the use of performance indicators by AQIS for inspection operations. The earliest official reference that we can find at the time of writing is a report by the Australian National Audit Office (ANAO, 2001). The ANAO pointed out (pages 77–83) that performance reporting had improved since 1997. AFFA (as it then was) reported effectiveness in two ways: by *volume*, the number of units upon which a problem was detected, and (for some programs) by *leakage*, the number or rate of units of concern following intervention. While these statistics were useful, the ANAO considered (p. 83) that they did not answer the question “how good is intervention at finding what it is meant to find?” The ANAO suggested a statistic, *seizure rate*, which is the proportion of units that should be seized that is actually seized. This is one measure of how good the inspectorate is at intercepting quarantine contamination; if the seizure rate is high, then only a small proportion of the risky material that approaches the country is entering the country. DAFF implemented the recommendation but used the name *effectiveness* instead, because the phrase seizure rate was already used to represent the proportion of approaching units that were seized.

The *effectiveness* of intervention is simply the proportion of non-compliant units that is actually detected by the intervention. Effectiveness is a very useful concept. Intervention might consist of a sequence of processes, not all of which might be necessary for a particular unit. For example electronic documents may be sufficient to conclude that a consignment is not a risk and the consignment can be released. On the other hand, electronic documents might indicate that that the consignment requires inspection or that the physical copies of the documents themselves need to be inspected before a decision can be made about release. The effectiveness of each component of intervention at doing what it is meant to do can be estimated by suitable leakage surveys. Furthermore, the overall effectiveness of the intervention can be estimated and interpreted as a performance metric.

The release of the ANAO report coincided with the Increased Quarantine Intervention (IQI) funding that was provided by the Commonwealth in response to the devastating foot and mouth disease outbreak in the United Kingdom. The IQI initiative set an intervention rate of 100% for a number of programs including some, such as the external surfaces of containers, which had not regularly been inspected in the past. The IQI initiative also set targets for the effectiveness at which both higher risk and risk material should be detected. Because there was generally 100% intervention, the effectiveness of intervention was the same as the effectiveness of inspection.

However, if fewer that 100% of units are inspected, then it is important to be precise about what "effectiveness" we are talking about: the effectiveness of a particular inspection process, or the effectiveness of all the intervention procedures taken as a whole, which may include screening, profiling, or monitoring of the pathway because it is known to be low risk.

We have already mentioned the need to relate the level of intervention to risk, which AQIS considers as the product of probability and consequences. This means that the type of intervention done need not necessarily be the same for each unit arriving in a pathway. Profiling provides a way to partially overcome this problem. Profiling involves using past inspection experience to divide the arriving units into a number of pathways according to their risk, and then concentrating the intervention effort on the more risky pathways. This exercise is complicated by the fact that the pathways can be different: for example, in the international passengers pathway the goal is to intercept risky goods, whereas in cargo the risk status of the goods are known in advance, and the goal is to intercept contamination.

In some cases AQIS is limited by the options for inspection: it either inspects or it releases; there is no half way point. In such a case AQIS would inspect all of the higher risk pathways and none of the lower risk pathways. However, as outlined in ACERA report 0804 (Robinson et al., 2008), the inspectorate also needs to ensure that it monitors the lower-risk pathways to ensure that they have not changed.

In other situations AQIS can adjust the effectiveness of the inspection that is performed, for example by adjusting the time that is spent on a particular activity. Again, profiling based on previous experience would be used to classify the arriving units into different pathways.

3

How Some Other Organizations Measure Performance

This chapter reviews performance indicators that are used by other inspectorates and regulators. We preface this review with the following observations.

We were unable to determine a common performance indicator across all the organizations that had similar missions. Furthermore, a number of inspectorates do not report their performance as inspectorates, for a range of reasons. Several are in the identical process of developing suitable indicators, and expressed great interest in the outcome of this report. Others do not collect data that would permit such reporting.

The range of functions undertaken by inspectorates is very broad. Many inspectorates that we reviewed operate in circumstances that are quite different to AQIS. For example, as noted earlier,

- AQIS can obtain a list of all activities on regulated pathways. The ACCC can obtain such a list for only a subset of its operations. EPA Victoria has regulatory responsibility for the collective outputs of organizations that it does not license.
- AQIS can represent its regulatory environment as a collection of pathways. Victoria Police has a much wider array of responsibilities under its remit, only some of which can be represented as pathways.

We interviewed personnel from a range of international, national, and state organizations that have regulatory responsibilities. The interviews were conducted by email, by telephone, and by personal visit, where possible. We also reviewed reporting mechanisms by some regulators, providing a summary of the inspection-related statistics that the organization reports.

The organisations that we approached are listed below. We copied each organization's description of its function from its website and summarised its relevance to this project.

3.1 Indicators

We now present a brief overview of the indicators that are reported by the organizations. We divide the indicators loosely into three main groups: how much effort is made, how well the inspectorate performs its responsibilities, and what effect the inspectorate has on the outcomes within its purview. The effect on outcomes can be measured in two complementary themes: the amount of risk reduced, and the amount of risk remaining.

Table 3.1: Summary of different types of performance indicators used by different organizations. The column names are explained in the following sections, but broadly speaking, *Effort* is a measure of resources expended, *Execution* is some measure of the quality of the resource usage, *Effect* is some measure of the impact on outcomes within the organization's scope, and *End Result* is some measure of the outcomes afterwards. In the table, *E* means that measures are reported externally and *I* that measures are reported internally. A blank means that we could find no evidence of the use of such measures. *U* means under consideration.

Organization	Effort	Execution	Effect	End Result
ACCC	E			
AQIS	E	I	E	
Customs	E		E	
EPA Victoria	E			
NZ MAF	E	I	E	E
US Customs	E	E	E	E
USDA APHIS	E	E	E	E

Ideally an inspectorate would use measures from both of these themes, but doing so may be prohibitively expensive. Within each group, we also note whether or not the indicator is published internally or externally.

3.1.1 Amount of Effort

The amount of effort that the inspectorate undertakes is a popular indicator of performance. Some measures of effort are under the control of the organization; for example an inspectorate may decide to perform a specific number of audits in a given time period. Others will vary due to influences that are outside the inspectorate's control, for example, the number of international passengers that arrive in a given year.

All of the organizations that we surveyed publicly reported indicators of the amount of effort, examples of which follow. Both the US and the Australian Customs organizations report the number of international passengers processed, as do MAF and APHIS. EPA Victoria reports the number of compliance inspections or audits of licenced sites, and ACCC reports the number of proceedings for cartel conduct that were instituted. The list for each organization is not exhaustive.

A concern with amount of effort as a performance indicator is that, in isolation, it presents a narrow and possibly misleading portion of the entire performance. From a regulatory point of view, there are several conflicting reasons that the amount of effort could fluctuate. For example, a reduction in the number of proceedings for cartel conduct instituted by the ACCC could signal that there are fewer instances of such conduct, so that the regulator is being effective in deterring the illegal behaviour, or that the instances are better hidden, or that the regulator is making less effort to find them. However, the amount of effort is easy for the regulator to measure, and sometimes appears in isolation.

3.1.2 Quality of Execution

Here, the quality of execution refers to how well the inspectorate carries out its role as an inspectorate. This quality can be thought of from two perspectives: how well the inspectorate chooses the units to be inspected, and how well the inspectorate carries out those inspections. In AQIS parlance, the first consideration is in the quality of profiling. How well do the profiles capture patterns of non-compliance, and therefore how well can AQIS distinguish between low- and high-risk units before inspection. The second consideration is in the effectiveness of inspection, that is, what is the probability that a non-compliant unit will be detected?

Public reporting of the quality of execution is rare. However, AQIS reported inspection effectiveness for international mail and passengers in 2008–9 (Tables 11 and 12 respectively). Another notable exception is the US Customs, quoted in Sparrow (2000, p. 66), which reported both the interception rate for vehicular inspection in 1996 (it was 19%; this is the pathway-level effectiveness) and the ratio with which profiled vehicles were non-compliant compared with random vehicles (it was 9; this is a measure of the quality of the profiling). We have found no other evidence of public reporting of these kinds of quantities.

Internal reporting of such indicators of quality was more common. For example, AQIS, and MAF report inspection effectiveness, as defined above. APHIS collects data that would be suitable for computing effectivness but it is unclear if they report it.

Measures of quality of execution are useful to an inspectorate because they provide a means to assess the utility, even the economic value, of intelligence about the pathways. There are also measures to assess the efficacy of inspections as a means of intercepting non-compliance. A point of attention with the use of measures of quality of execution as a performance indicator is that they are typically estimated using data from dedicated small-scale surveys, and are therefore subject to sample-based uncertainty. This means that the estimates of the measures can fluctuate considerably, which may create the impression that the underlying processes are changing more than they really are. Therefore short-term interpretation of such figures needs to be approached carefully.

3.1.3 Effect on Outcomes

As noted above, the effect of the inspectorate on outcomes can be measured both by the amount of risk that is removed, which is covered in this section, and the amount of risk remaining, in the next section.

Generally speaking, the amount of risk reduced is easier to measure. For example, inspectorates can report the number of harms intercepted, as is done by both Australian and US Customs. AQIS reports the number of mail items seized, and the number of items seized from passemgers. MAF reports the number of sea containers found with actionable biosecurity risk material.

3.1.4 End Result

The amount of risk remaining is more challenging because it requires information about the leakage of risk. This information is generally only available from leakage surveys and the like, as mentioned in the previous section. An example of such a measure is post-intervention compliance, which reports the estimated rate at which units of the pathway are compliant with regulation after the inspectorate's intervention. This measure is publicly reported for international air passengers by APHIS and MAF. As an indirect example, Customs uses a suite of indicators to try to assess the effect of its operations on illegal drug availability: needle exchange , ambulance call out frequency, household surveys, and so on.

3.2 Review of Organizations with Similar Challenges

3.2.1 Quarantine Operations

New Zealand Ministry of Agriculture and Forestry

> *Our mission is to enhance New Zealand's natural advantage. We do this by: encouraging high-performing sectors; developing safe and freer trade; ensuring healthy New Zealanders; and by protecting our natural resources for the benefit of future generations.*

The following information was provided by Victoria Allison and Carolyn Whyte of MAF.

MAF is also working on a risk-return approach, but taking a different direction to AQIS in some areas. Effectiveness is one of several performance metrics. MAF tries to estimate the quantum of "risk" arriving in a pathway as well, which provides a rough idea of whether the inspection effectiveness is acceptable. For instance, if MAF determines a risk value of 300 per month due to leakage on pathway A, and a risk value of 30,000 per month due to leakage on pathway B, the fact that B is 90% effective and A is only 20% effective doesn't imply that MAF should put more resources into A. For this reason, compliance targets are likely to be set higher for higher volume pathways.

As an example, in the relatively high volume passenger pathway, the compliance target is 98.5%: at least 98.5% of passengers must be free of all biosecurity risk at the time they leave the airport. These kinds of compliance measures do not take in to account different levels of risk, and thus do not account for targeting of resources towards passengers likely to carry higher-risk units. To address this issue, MAF calculates two values: one overall, and a second that only includes medium and high-risk leakage values.

In the passenger and mail pathways, MAF uses a risk-unit system to track changes through time, and to compare across work-sites[1]. The risk-unit system involves the quantification of risk for non-compliance.

The concept of risk units does not readily translate to commercial shipments of risk goods because:

- The scale was developed for relatively small quantities of risk taken from private consignments, and it is not known how risk scales with quantity over orders of magnitude;
- Risk for private consignments is based on products that have not undergone some kind of certification, where the product may carry a pest or disease. For commercially certified consignments, only units known to be infested or contaminated are a problem. Thus, it is difficult to compare an apple taken from a passenger with a commercial consignment of apples where regulated pests have been identified.

---

[1] Not to be confused with inspection units, used within this report to refer to consignments, passengers, and the like, to which intervention is applied.

- Risk units are difficult to apply to vessels or conveyances that are contaminated with hitch-hikers.

Because of these issues, MAF does not attempt to use risk units to measure risk in the cargo and craft/vessel pathways. Within the cargo pathway, the prescribed level of intervention depends on the relative risk of the commodity and the compliance history for that commodity, importer, or exporter.

The initial approach developed for the cargo and craft pathways was conceived as applying to a whole segment of goods, and focused exclusively on the inspection regime in place at the border. MAF now believes this approach is too simplistic, and is moving towards identifying optimal points of intervention for each segment (or in some cases sub-segment), with audit or verification schemes developed to measure if compliance is achieved at these points.

MAF is now starting to think about future directions. MAF is working with Customs in New Zealand to develop a new border database. This database will depend on a 'risk rating' for each arriving consignment. The risk rating will be based on the likelihood of association with different pests, and the severity of the pest. The risk ranking will also factor in past compliance on the pathway, results of audits on off-shore systems, etc. The risk ranking would be used to determine audit frequency (low through to 100% of all arriving consignments), and also contribute to decision making about how high a priority any one inspection should be. Consignments with a high risk ranking would have to be inspected every time, while those with a low risk-ranking might have inspections deferred if resources are lacking.

Since the earlier drafts of this report, MAF has begun measuring compliance rate on passengers and mail pathways, and reporting this quentity externally. MAF has a formal compliance target in place (98.5%), and this is used as part of decision making about how to use resources within the pathway, and when to redirect to other activities.

3.2.2 USDA Animal and Plant Health Inspection Service

> *"Protecting American agriculture" is the basic charge of the U.S. Department of Agriculture's (USDA) Animal and Plant Health Inspection Service (APHIS). APHIS provides leadership in ensuring the health and care of animals and plants. The agency improves agricultural productivity and competitiveness and contributes to the national economy and the public health.*

The following information was provided by Michael Caporaletti, Program Analyst United States Department of Agriculture Animal and Plant Health Inspection Service (APHIS).

In the mid 1990s, Congress passed the Government Performance and Results Act (GPRA). Prior to the passage of the GPRA, most agencies only reported outputs or work counts for federally funded programs. The GPRA required development of indicators to measure program results.

In response to the GPRA, Plant Protection and Quarantine (PPQ) implemented Agriculture Quarantine Inspection Monitoring (AQIM) activities for various pathways into the United States to assess the performance of the Agricultural Quarantine and Inspection (AQI) program at ports of entry. AQIM activities at the ports of entry consist of daily or weekly random sampling of passenger baggage, vehicles, mail or cargo that provide additional information (data) on the potential agricultural risks that are approaching US ports of entry. These data can substantiate known potential risk pathways or help identify additional trends, patterns or areas of potential risk to agriculture. This could then lead to improvements in a port's agricultural risk management decisions, improved selection criteria and better use of a port's current and future agriculture resources. The AQIM data are also used in the annual performance measures that USDA reports to Congress.

Here is a brief explanation of how the approach is deployed in the air passenger environment. The outcome that PPQ measures is the percentage of passengers in compliance with agriculture regulations[2]. AQIM is in place at 25 airports. Approximately 90 percent of all international passengers arriving in the US enter at one of these 25 airports. A total of 300 random samples is selected at each location each month. Results of the random inspections are used to estimate the percentage of arriving passengers (approach rate) with a prohibited agricultural commodity. Hypothetically, if the data indicate that five percent of passengers have a prohibited unit (QMI) then an estimated 95 percent do not have a QMI and are in compliance with agricultural regulations.

PPQ then looks at the actual results recorded in the Work Accomplishment Data System (WADS). The WADS data show how many QMIs were seized during inspection of passenger baggage. A one to one ratio for passengers to QMIs is assumed. This is not always the case as some passengers do have more than one QMI. If 10,000 QMIs were seized from a total of 1,000,000 arriving passengers then an estimated one percent of the passengers would be brought into compliance as a result of an agriculture inspection. This total, when added to the estimated 95 percent in compliance, results in a total estimated percentage of passengers in compliance with agriculture regulations of 96 percent.

PPQ uses several other measures in conjunction with the compliance measure, for example, program outputs such as the number of arriving passengers, the number of QMIs seized, and the number of passengers referred to agriculture secondary for inspection. The latter inspection involves all baggage being x-rayed, and based on x-ray, some baggage will be opened and physically inspected. Finally, PPQ has a measure of program efficiency. A measure compares how much better targeting of passengers is when compared to random sampling.

In addition to the above described measures for the air passenger environment, there are similar measures in place for vehicles arriving on the Northern and Southern Border. AQIM is also used to estimate the approach rate in the mail, pedestrian, cruise ship and cargo (air, land, and sea) pathways. However, a compliance goal for these pathways has not been established.

Here is an example of how the targeting is compared to random inspection.

- Random — This is taken directly from AQIM sampling results. Using the earlier example, assume that 5 percent of the passengers sampled had a prohibited unit.
- Targeting — To calculate this rate, use the WADS data. For example, assume WADS indicates that 800,000 prohibited units were seized and a total of 8,000,000 passengers were referred for agricultural inspection. 800,000/8,000,000 = a rate of 0.10 (or 10 percent).
- The percentage of positives for passengers targeted (10 percent) is two times greater than the percentage of passengers randomly selected (5 percent).

The data reported by APHIS is collected in the ports of entry and reported to APHIS–PPQ by Customs and Border Protection Agriculture Specialists.

---

[2] We will refer to this quantity as post-intervention compliance (PIC) in this report.

3.2.3 Other Inspectorates

Australian Customs and Border Protection Service

> *Australian Customs and Border Protection Service manages the security and integrity of Australia's borders. It works closely with other government and international agencies, in particular the Australian Federal Police, the Australian Quarantine and Inspection Service, the Department of Immigration and Citizenship and the Department of Defence, to detect and deter unlawful movement of goods and people across the border.*

Customs undertakes several types of activities at the border: detecting and deterring unlawful movement of goods and people; verifying the status of imports of legal but restricted items, such as firearms; and collecting revenue. Each of these activities demands different kinds of performance metrics.

Two types of performance measures are used, broadly divided into measures of effort and impact. An example of a measure of effort is in the importation of firearms: 100% of guns need to be stopped and checked, so the performance measure is a count. Customs also reports on rates of declared and undeclared units, and this is documented in the annual report. An example of a measure of impact is in the use of the Australian Federal Police harm index. For example, a press release might read: Customs stopped x kilos of cocaine, saving the community y. Also Customs will try to look at broader market, e.g., the street price of drugs will spike after a seizure. Internally, Customs will look at key risk indicators as triggers for action such as poor detection counts, poor press, and public perception.

Direct measurement of performance with regards to detecting and deterring criminal activities is a difficult task. An example of an indirect metric is an estimate of the amount of heroin available on the street. This metric is useful for the following two reasons. First, heroin is unequivocally imported. Second, the incidence of heroin is relatively easy to measure (in comparison to other illicit substances) using a range of metrics; for example, using household surveys, drug use monitoring in police lock-ups, ambulance call out rates, needle exchange, and deaths by drug overdose.

Customs also establishes partnerships with other organizations and carries out regulatory activities on behalf of those organizations. For example, the ACCC seeks to prevent the importation of certain goods of consumer safety concern, such as novelty cigarette lighters or cosmetics that contain heavy metals. Customs may carry out inspections of imports on behalf of the ACCC and other partners.

There are about 45 policy agencies and over 70 categories of restricted or prohibited goods, ranging from products containing cat and dog fur to weapons of mass destruction. Customs takes an intelligence-led, risk-based approach to detecting illicit importation of these goods. Intervention activity is based on risk, and ranges from actively seeking to disrupt and prosecute the illicit importation of goods that represent the highest level of risk to national security and/or community health and safety (such as illicit firearms, narcotics and child pornography), through to monitoring for goods that are regulated, but which do not present an immediate threat to the community (such as inefficient incandescent light globes).

The challenges faced by this arm of Customs are similar to those faced by some areas of AQIS: there is a pathway of goods, some of which are prohibited, and others of which are regulated and therefore require appropriate paperwork. The organization is currently in the process of developing new ways of conceptualizing the space in which they are operating, particularly as it relates to what represents an appropriate, risk-based approach to intervening in respect of goods that present a lower risk. This work is also examining ways of measuring and reporting on performance. It will likely depend on the nature of the risks faced, and will be developed in consultation with partners: some partners may only require exchange of information to support domestic regulatory activity (such as is proposed for products made from or containing illegally logged timber), while others may require seizures (some illicit weapons, such as daggers) or prosecutions (child pornography). The outcome of the present report may be relevant.

We are grateful to four Australian Customs and Border Protection Service (Customs) officers whom we interviewed for this section.

United States Customs and Border Protection

> *U.S. Customs and Border Protection secures the homeland by preventing the illegal entry of people and goods while facilitating legitimate travel and trade.*

In Sparrow (2000, p. 66), John Hill, the director of the the Customs Service development of performance measurement techniques, summarized performance for vehicular inspection in FY 1996 by reporting

1. the estimated approach rate of contamination (Significant violations occurred in about 5 out of every 100,000 vehicles.)
2. the estimated pathway-level effectiveness (The CS apprehend about 19% of those violations), and
3. the utility of profiling in this context (The inspector's selection is about 9 times better than purely random choice.)

These statistics reflect the inherent state of the pathway, the utility of inspections for detecting non-compliance, and the utility of screening or profiling as an intervention tool.

3.3 Review of Organizations with Different Challenges

3.3.1 Australian Competition and Consumer Commission

> *The ACCC promotes competition and fair trade in the market place to benefit consumers, businesses and the community. It also regulates national infrastructure services. Its primary responsibility is to ensure that individuals and businesses comply with the Commonwealth competition, fair trading and consumer protection laws.*

The following information about Australian Competition and Consumer Commission (ACCC) was provided by Steve Hutchison, DGM Product Safety (ACCC).

Among its other activities, the ACCC administers 59 mandatory standards or bans for products ranging from children's toys to bunkbeds and car jacks. Injury data are used as a source of risk information. The substantial challenge that ACCC faces in this area is in identifying the cohorts for monitoring. Monitoring is possible for some products — e.g., bunkbeds — and very difficult for others, e.g., novelty toys, because it is difficult to obtain a definitive list of suppliers.

Various international bodies are trying to develop risk models. Development of performance measures by ACCC is still underway. The outcome of the present report may be of interest.

3.3.2 Victorian Environmental Protection Agency

EPA Victoria's purpose is to protect, care for and improve our environment.

The following information about Victorian Environmental Protection Agency (EPA Victoria) was provided by Kari Sann, Corporate Strategy Adviser, EPA Victoria.

EPA Victoria is presently changing in response to a new executive, including the adoption of outcome mapping (logic maps), etc. The organization presently uses a scorecard approach, with activity and outcome-based performance indicators such as

- Less than 30,000T of High-Hazard Waste disposed to landfill from manufacturing sources
- 200 compliance inspections or audits of licensed sites are conducted focusing on high risk sectors
- 100% increase in the number of notices issued relating to the illegal dumping of waste

The EPA's 2010/11 Business Plan also identifies some performance indicators, such as: Increase in EPA notices issued for illegal dumping of waste: 15%.

The challenges faced by EPA Victoria and similar regulators is that it is difficult to identify outcomes that are related to the regulator's activities, hence the focus in performance is more heavily weighted towards effort.

EPA is currently finalising a 5 year Corporate Plan. The Plan is underpinned by logic maps that link annual activity to 20 year environmental outcomes. The scorecard is being updated to reflect measures at activity, interim outcome and long term outcome levels.

4

Post-Intervention Compliance

4.1 Introduction

In this chapter, we expand on Post-Intervention Compliance (PIC) which was briefly introduced in Section 3.2.2, focusing on how it could be best measured and used in the context of AQIS's existing data holdings and current practice. PIC is defined as the estimated percentage (or proportion) of inspection units on the pathway that comply with quarantine regulations after AQIS intervention. The percentage is calculated relative to the number of units arriving on the pathway, rather than the number of units on the pathway after intervention.

PIC can be used in two ways: first, to *report* the past compliance performance of a specific pathway, or perhaps a trend in time, and second, to *predict* the future compliance performance of a specific pathway based on some historic inspection data and a plan for future intervention. The PIC can thus be used both as a reporting tool and as a planning tool to assist the pathway manager in choosing intervention strategies and setting intervention levels. The PIC may also be compared with the *before-intervention compliance* (BIC), which is one minus the approach rate.

As we show below, the calculation of PIC involves the calculation of several other statistics that are also useful as measures of certain aspects of performance. These other measures can be used to assess different components of inspectorate performance, and to guide strategic investment of resources.

The fundamental components of PIC are (see Table 4.1):

- pathway volume (number of units arriving),
- inspection effort (number of units inspected),
- pathway approach count (number of units arriving that are not compliant), and
- inspection effectiveness (probability that inspection will detect existing non-compliance associated with a unit).

Unlike effectiveness, which measures the ability of an intervention process to detect non-compliance, the PIC is related to the level of non-compliance after an intervention process.

To achieve a desired PIC value, one uses an estimate of the approach rate (the rate of non-compliance before intervention, computed using the approach count and the volume) to determine what level of inspection would be required to achieve the desired level of compliance. Thereafter one would monitor the PIC and associated statistics to see if

Table 4.1: Components of the Post-Intervention Compliance (PIC) performance indicator.

Measure	Definition	Interpretation
Approach Rate	Proportion of units on the pathway that are not compliant	Pathway cleanliness
Inspection Count	Number of units on the pathway that are inspected	Inspectorate effort
Inspection Effectiveness	The probability that a non-compliant unit that is inspected will be intercepted.	Inspection quality
Leakage Count	Number of units that would or should have been intercepted by intervention but were not	Pathway risk
Volume	Number of units on the pathway	Pathway workload

the approach rate has changed. For example, the approach rate may have decreased because of an extensive advertising campaign, so that less intensive (and therefore cheaper) intervention might be appropriate. Alternatively, the approach rate may have increased, which might lead to a greater level of intervention being required. The rate of monitoring would depend on characteristics of the pathway.

A potentially useful extension to the PIC that could be used to guide the rate of monitoring is the change in PIC between two times (δ-PIC) with a confidence interval on the change. The change in the PIC would provide information about the trend, and the confidence interval would give a statistical indication of how conclusive the information about the trend is. If the confidence interval on the difference between successive months is too wide then information about the monthly trend is too diluted to be useful, and decisions could be made on a quarterly or even longer basis.

4.2 Interpretation

The PIC and its components have concrete interpretations that can be easily communicated and reflect directly on important aspects of quarantine performance (see Table 4.1). These characteristics are advantageous for a performance indicator. As provided in the example, the PIC is based on measures of

- the volume of units on the pathway, which indicates the substance of the management task for AQIS,
- the number of units inspected, which indicates the amount of effort undertaken by AQIS with regard to this pathway, if that is different from the volume,
- the approach rate of non-compliance on the pathway, which is one component of the biosecurity risk represented by the pathway,
- the effectiveness of inspections, which is useful to know for the purposes of training and inspection resources deployment, and
- the pathway leakage, which is an estimate of the number of units on the pathway that should or would have been intercepted by the intervention. Some of these units may be non-compliant.

The specific details concerning the implementation and interpretation of the PIC are the subject of ongoing work in ACERA Project 1101 B “Performance Indicators”.

The PIC focuses directly on the outcome that the regulator seeks: compliance. PIC reflects the end state of the pathway, and combines characteristics of the pathway with measures of the performance of the intervention. A high inspection rate combined with a high inspection effectiveness may be sufficient for the pathway to be compliant but they may not be necessary; if the compliance is already high then all that is required is a reliable measure of compliance that can be reported.

Compliance on arrival ($1 -$ approach rate) can be used to justify different intervention levels or intensities, and gives stakeholders involved in the pathway something to aim for — if they increase compliance to x, then they may experience fewer interventions on arrival.

4.3 Calculation of Post-Intervention Compliance

The post-intervention compliance (PIC) is calculated as

$$\text{PIC} = \frac{v - \hat{L}}{v} \tag{4.1}$$

where v is the number of units on the pathway, also called the volume of the pathway, and $\hat{L}$ is the estimated number of units in the pathway that are still non-compliant after intervention[1].

Note that some intervention units may be associated with more than one instance of non-compliance; a typical example is air passengers, who may carry multiple items of biosecurity risk material. Such units should still only count as one non-compliant unit; hence the measure represents quarantine compliance for the pathway.

As the PIC is estimated from sample data, among other kinds, it is subject to sample-based uncertainty, among other kinds. Sample-based uncertainty is usually reported for statistics using a *confidence interval.* With 100% inspection, confidence limits for leakage are straightforward to calculate. However, if only a proportion of the units are inspected, the computation of confidence intervals is complicated by the fact that the number of non-conformities found would vary depending on which units are inspected.

While the estimation for the overall PIC from multiple pathways is a reasonably easy extension of the standard method for calculating PIC for a single pathway, calculating confidence limits is more complicated. Several expensive plug-ins provide this functionality to Microsoft Excel, for example, @Risk and Risk Solver. Alternatively, the intervals can also be computed in the free statistical environment, R (R Development Core Team, 2010); for an example see Appendix B. It is also possible to program Excel to compute these confidence limits.

4.4 Estimating the Leakage

The simplest way to estimate the leakage ($\hat{L}$) is to do a leakage survey of all *compliant* post-intervention units in which each unit has an equal chance of being sampled, and then to analyze the resulting data as described in this section. Generally, a ‘leakage survey’ refers

---

[1] In this report, the *hat* will be used to signify quantities that are estimated. Here, we assume that the pathway volume is known (or can be known) but the leakage must be estimated.

to a sample of consignments that is taken after the usual quarantine intervention, and is used to estimate various statistics about the leakage of BRM through the intervention. For example, in the quarantine x-ray inspection of reportable documents, a randomly-selected proportion of the units is directed through the x-ray machine twice, using different orientations, and a randomly-selected proportion of the units so directed is then opened and physically inspected. This re-inspection process is a leakge survey.

The leakage survey, or its equivalent, is a key component of estimating the performance indicator. Estimates of the PIC in pathways that lack a leakage survey will rely on the assumption that all non-compliance in inspected units is detected. If a leakage survey is not in place then the inspection effectiveness will have to be guessed, for example by expert opinion.

For the purposes of this document, we will consider the leakage survey to be an intrinsic and regular element of the intervention: it is the final component of intervention before units are released. Therefore, non-compliant units that are detected by the leakage survey are considered to be detected by the “total intervention process”. This assumption is the subject of some disagreement among the authors with some feeling that when calculating the PIC, the leakage survey should not be considered part of the total intervention process because units found by it would otherwise have been leaked. However, we all acknowledge that the practical difference will be negligible. Where the alternative assumption makes a difference to the development of the equations, we will note it. There is further discussion on the estimation of leakage in Section 5.5.

We will also assume that the leakage surveys are 100% effective, that is, they detect all non-compliance in the leakage survey units. If the effectiveness of the leakage survey is known, or can be estimated, then it can be used to adjust the following formulas. This assumption is used to simplify the following exposition and is usually considered defensible, even though it is not possible. In general, it is preferable for a leakage survey to have higher effectiveness than the regular survey, so inspections of units in the leakage survey will often be more detailed.

The way to estimate the leakage depends on the design of the inspection regime. Here we cover two common scenarios:

1. The pathway is completely inspected and a leakage survey is performed, for example the inspection of reportable documents during IQI.
2. The pathway is sampled and a leakage survey is performed, for example, the inspection of reportable documents in a risk-based approach.

The estimation also depends on how the intercepted units are handled. In many cases, the non-compliance will be corrected and the unit will be released. We will refer to these inspections as *rectifying*. In other cases, the unit may be destroyed, or returned. Such inspections are referred to as *non-rectifying*.

Also, some leakage surveys include only those units that were cleared by inspection, excluding those units that have been intercepted and rectified.

Numerous variations on these themes are presently in use by AQIS. Each of these differences leads to a small change in the development of the leakage estimate. Examples include

- For passengers in particular, but also on other pathways, any BRM found is either destroyed or treated and the passenger then proceeds, possibly subject to a leakage survey. In such a case, every inspected passenger has a chance of being included in the leakage survey.

- However, on other pathways, units found with BRM will not be subject to a leakage survey either because it is considered unnecessary because of the thoroughness of the inspection or because the unit no longer exists (destroyed or re-exported).
- There will also be situations where something in between happens. Some risk material is destroyed or exported while other material is rectified. An example of this is in the importation of perishable goods; the importer can choose to have a contaminated consignment treated, destroyed, or re-exported.

Covering all AQIS scenarios is beyond the scope of this report, rather, we try to present the simplest and most general versions. A general development that can be adapted to other contexts is provided in Appendix A.

4.4.1 Complete Inspection with a Leakage Survey.

The first example is for the case in which the pathway is completely inspected and a leakage survey is performed. We assume that the leakage survey is a sub-sample of all inspected units after any rectifying actions have been done to non-compliant units. This assumption is reasonable for, e.g., the inspection of air passengers, and probably reasonable for reportable documents. See Figure 4.1 for a flowchart that presents this inspection process.

We define

v as the volume of the pathway (number of intervention units), each of which is inspected in this example,

b as the number of inspected units that are found to contain biosecurity risk material (BRM) (not including those found by the leakage survey),

n as the number of units inspected in the leakage survey,

y as the number of units inspected in the leakage survey that are found to be not compliant,

l as the leakage count of units *that were inspected*,

L as the pathway-level leakage count, that is, the number of contaminated consignments that are not intercepted,

a as the approach count,

a_r as the approach rate, and

e as the inspection effectiveness, defined as the proportion of non-conforming units detected among all those non-conforming units that are inspected.

The estimated leakage count from the inspection is the number of inspected units, multiplied by the non-compliance rate as computed from the leakage survey.

$$\hat{l} = \frac{v \times y}{n} \tag{4.2}$$

The estimated approach count, $\hat{a}$, is the number of units detected during the inspection with BRM, added to the estimated number of missed units,

$$\hat{a} = b + \hat{l} \tag{4.3}$$

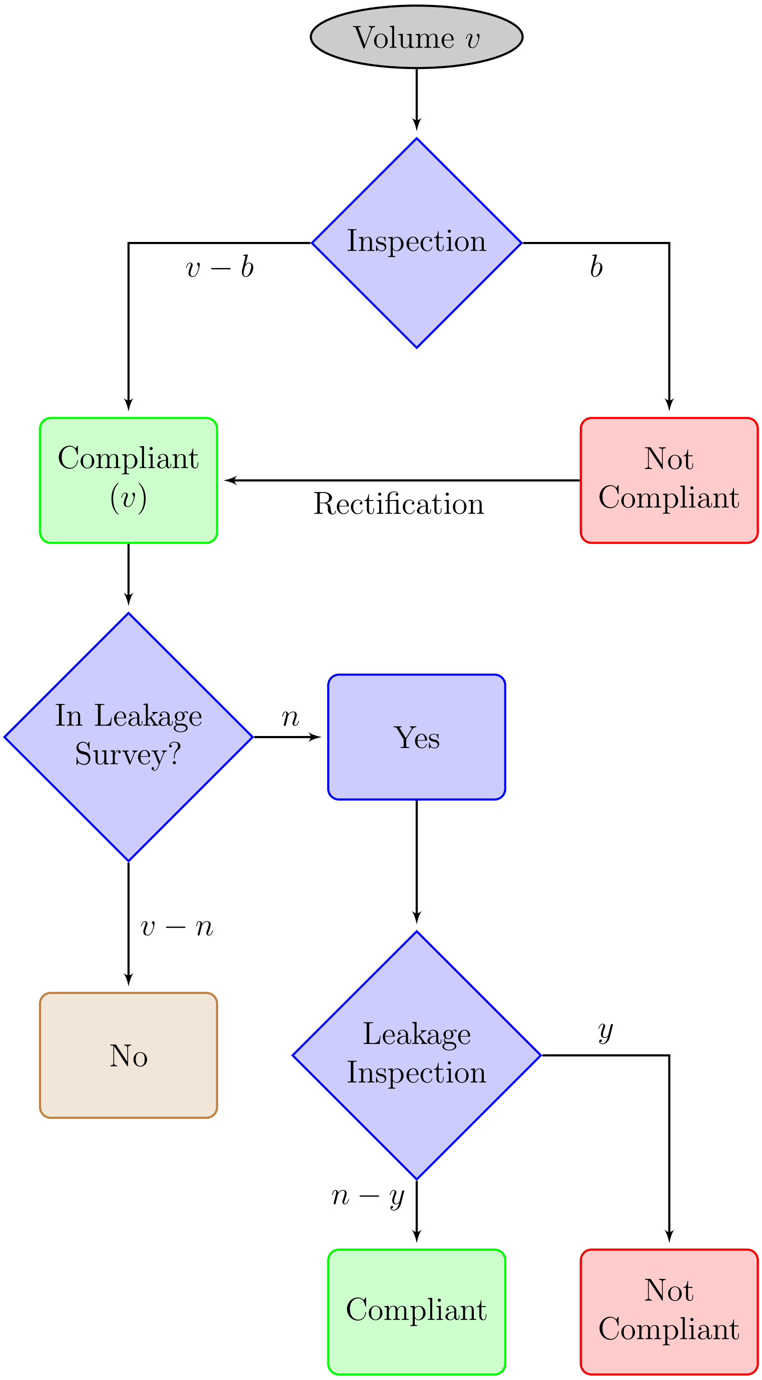

Figure 4.1: Flow chart for complete inspection of pathway with leakage survey.

Then the estimated effectiveness of inspection, $\hat{e}$, is the ratio of the number of units with BRM detected during the inspection and the estimated number of units that had BRM.

$$\hat{e} = \frac{b}{\hat{a}} \tag{4.4}$$

Finally, the pathway leakage is the estimated number of units with BRM minus the intercepted number of units with BRM.

$$\hat{L} = \hat{a} - b - y = \hat{l} - y \tag{4.5}$$

which can also be written as

$$\hat{L} = y\left(\frac{v}{n} - 1\right) \tag{4.6}$$

Note that these equations reflect the assumption that the leakage survey is an intrinsic part of intervention, meaning that units intercepted by the leakage survey but not the original inspection are not themsleves counted as being leaked, even though they are used to calculate the leakage. Therefore the y units intercepted by the leakage survey should not be included in the predicted leakage. If this assumption does not hold, then we would use

$$\hat{L} = \frac{y \times v}{n} \tag{4.7}$$

Example

We use a simple numerical example to demonstrate the PIC as a tool for (i) reporting performance and (ii) planning future intervention. We present the best estimates of the quantities of interest; confidence intervals are deferred to Appendix B.

The examples have been cut from R output, and look something like this:

```
> (l_hat = v * (y/n))

[1] 66.66667
```

The angle bracket `>` is R's way of asking for something to do, as is the `+` sign. Here, we have asked R to calculate the inspection level leakage, using equation 4.2. R calculates the result and give us the answer after a line number (here, [1]).

Example of Reporting Performance Assume that the following values have resulted from historical inspection data. We emphasize that these statistics are purely for the purposes of providing an example, and bear no necessary relation to any inspection data.

```
> v = 10000
> b = 20
> n = 300
> y = 2
```

Then the estimated inspection leakage (as a count of non-compliant units) is

```
> (l_hat = v * (y/n))
```

```
[1] 66.66667
```

the estimated approach count is

```
> (a_hat = b + l_hat)
```

```
[1] 86.66667
```

and the estimated inspection effectiveness (reported as a percentage) is

```
> (e_hat = b/a_hat) * 100
```

```
[1] 23.07692
```

This is the percentage of non-compliant units inspected in which non-compliance is actually detected. The estimated pathway leakage *count* is

```
> (L_hat = l_hat - y)
```

```
[1] 64.66667
```

or equivalently,

```
> (L_hat = y * (v/n - 1))
```

```
[1] 64.66667
```

Finally, if the leakage survey is not considered part of the intervention, we use

```
> y * v/n
```

```
[1] 66.66667
```

The PIC is

```
> (PIC = (v - L_hat)/v) * 100
```

```
[1] 99.35333
```

Several results are noteworthy.

First, the pathway has a high rate of compliance after intervention despite the relatively poor inspection effectiveness ($\hat{e} = 23.1\%$). The high compliance results from the inherent cleanliness of the pathway. This cleanliness can be summarized by the approach rate, expressed as a percentage of the volume, viz. 0.87%), or by the Before Intervention Compliance (BIC), which is the one minus the approach rate (here, 99.13%) The pathway cleanliness is affected by higher-level operations such as public relations campaigns and negotiating the definition of non-compliance using science, and is monitored as a fundamental function of pathway risk management. The cleanliness may also result from the fundamental nature of the pathway, in that the pathway may not be a suitable vector for non-compliance.

Second, as noted above, PIC has a concrete real-world interpretation, conditional on the assumptions laid out below. An estimated 99.35% of the pathway is compliant with quarantine regulations after intervention. Also, the components that are needed to compute it all have concrete interpretations, and each further illuminates the environment in which the inspectorate is operating.

We note in passing that we can also compute the estimated post-intervention leakage *rate* (here, expressed as a percentage) as

```
> L_hat/v * 100
[1] 0.6466667
```

and the before intervention compliance (BIC, here, expressed as a percentage) as

```
> (v - a_hat)/v * 100
[1] 99.13333
```

Example of Planning Future Intervention We now demonstrate how the PIC and its components can be used by a pathway manager to predict the outcomes of two different changes in the intervention regime: (1) reducing the approach rate, for example via a targeted public relations campaign, and (2) increasing inspection effectiveness, for example by training.

For convenience we will summarise the key results under the current intervention as obtained in the previous paragraphs:

PIC: 99.35 %
BIC: 99.13 %
Approach Rate ($\hat{a}_r$): 0.87 %
Inspection Effectiveness ($\hat{e}$): 23.08 %
Pathway Leakage ($\hat{L}$): 64.67 units per year
Inspection Leakage ($\hat{l}$): 66.67 units per year
Volume (v): 10000 units per year.

The pathway manager now has two distinct, but not mutually exclusive, ways to increase the future pathway compliance rate, each of which will be reflected in a component of the performance indicator.

1. *Reduce the approach rate.* We could, for example, halve the approach rate via a well-targeted public relations campaign. In the passengers pathway, for example, a PR campaign might be designed to increase the declaration rate of units that the passengers think might be BRM. There are others ways that the BIC could be reduced. For example, greater depth of identification might separate pests that are of no quarantine concern from similar-looking pests that are of quarantine concern. We shall assume for this example that we can reduce the approach rate by 50%. Then the PIC changes as follows:

   PIC: 99.68 %
   BIC: 99.57 %
   Approach Rate ($\hat{a}_r$): 0.43 %
   Inspection Effectiveness ($\hat{e}$): 23.08 %
   Pathway Leakage ($\hat{L}$): 32.33 units per year
   Inspection Leakage ($\hat{l}$): 33.33 units per year
   Volume (v): 10000 units per year.

2. *Increase inspection effectiveness.* We could, for example, approximately double the inspection-level effectiveness by further training, incentive schemes, improving inspection environment, etc. Then the inspection leakage reduces, but not quite to half of its previous level.

   PIC: 99.55 %
   BIC: 99.13 %
   Approach Rate ($\hat{a}_r$): 0.87 %
   Inspection Effectiveness ($\hat{e}$): 46.09 %
   Pathway Leakage ($\hat{L}$): 45.38 units per year
   Inspection Leakage ($\hat{l}$): 46.78 units per year
   Volume (v): 10000 units per year.

Such a calculation could be carried out using historical inspection data or based on a snapshot study that involved carefully-designed changes to the intervention strategy.

4.4.2 Sampled Inspection with a Leakage Survey.

The second example is for the case in which the pathway is sampled for inspection (ie. only a proportion of the units are inspected) and a leakage survey is performed on those units inspected. An example is the inspection of reportable documents during risk-return pathway management (Robinson et al., 2010). Sampling of inspection units might be used for pathways that are considered to be inherently low risk, for example because the approach rate is low, or the specific end use sharply reduces the probability of establishment and spread of invasives. The management goal for such a pathway would be monitoring the non-compliance rate to be confident that it remains sufficiently low. We again assume that the leakage survey is a subsample of the inspection units that were found to be compliant, and the inspections are rectifying. See Figure 4.2 for a flowchart of the inspection process. We define, mostly as above,

v as the volume of the pathway (number of intervention units),

i as the number of intervention units inspected,

b as the number of inspected units that are found to contain biosecurity risk material (BRM) (not including those in the leakage survey),

n as the number of units inspected in the leakage survey,

y as the number of units inspected in the leakage survey that have non-compliance,

l as the leakage count of units *that were inspected*,

L as the pathway-level leakage count,

a as the approach count,

a_r as the approach rate, and

e as the inspection effectiveness, defined as the proportion of non-conforming units detected among all those non-conforming units that are inspected.

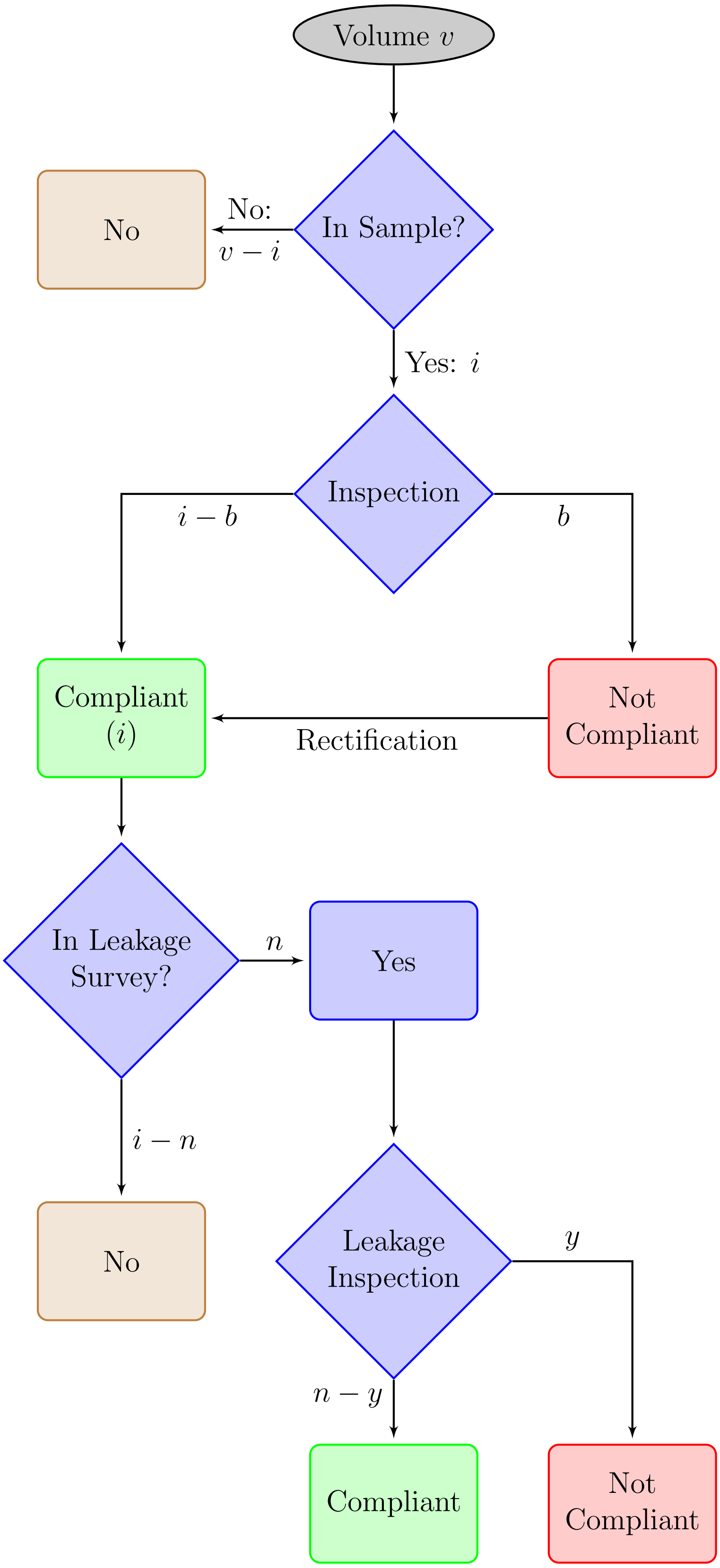

Figure 4.2: Flow chart for sampled inspection of pathway with leakage survey.

Now we have two contributions to pathway leakage: leakage through the inspection and leakage via the uninspected units. The inspection leakage count, l, is estimated by the number of units inspected that were found to be compliant.

$$\hat{l} = \frac{i \times y}{n} \tag{4.8}$$

The estimated inspection effectiveness $\hat{e}$ is then the ratio of what was found over the estimated non-compliance among the inspected units.

$$\hat{e} = \frac{b}{b + \hat{l}} \tag{4.9}$$

Then, the estimated approach count $\hat{a}$ for the pathway is the sum of the detected non-compliant and estimated undetected units, inflated to the population level.

$$\hat{a} = (b + \hat{l})\frac{v}{i} \tag{4.10}$$

Finally the estimated pathway leakage count is the difference between the estimated non-compliance and the detected non-compliance.

$$\hat{L} = \hat{a} - b - y \tag{4.11}$$

We can combine all these equations to determine $\hat{L}$ directly:

$$\hat{L} = b \times \left(\frac{v}{i} - 1\right) + y \times \left(\frac{v}{n} - 1\right) \tag{4.12}$$

The first term estimates the number of units with BRM that we would have found in the $(v - i)$ units we didn't inspect. The second term estimates the number of units with BRM that we would have missed because inspection is not perfect.

Again, these equations reflect the assumption that the leakage survey is an intrinsic part of intervention, and therefore that the y units intercepted by the leakage survey should not be included in the predicted leakage. If this assumption does not hold, then we would use

$$\hat{L} = b \times \left(\frac{v}{i} - 1\right) + y \times \left(\frac{v}{n}\right) \tag{4.13}$$

Example

We next provide a simple example with characteristics similar to the previous one. We emphasize that these statistics are purely for the purposes of providing an example, and bear no necessary relation to any inspection data. See page 26 for a brief explanation of the R output.

```
> v = 1e+05
> i = 10000
> b = 20
> n = 300
> y = 2
```

Example of Reporting Performance Then the estimated inspection leakage is

```
> (l_hat = i * (y/n))

[1] 66.66667
```

and the estimated inspection effectiveness (reported as a percentage) is

```
> (e_hat = b/(b + l_hat)) * 100

[1] 23.07692
```

so the estimated approach count is

```
> (a_hat = (b + l_hat) * v/i)

[1] 866.6667
```

and so the estimated pathway leakage *count* is

```
> (L_hat = a_hat - b - y)

[1] 844.6667
```

or

```
> (L_hat = b * (v/i - 1) + y * ((v/n) - 1))

[1] 844.6667
```

or, if the leakage survey is not part of the intervention, use

```
> b * (v/i - 1) + y * ((v/n))

[1] 846.6667
```

and the PIC is

```
> (PIC = (v - L_hat)/v) * 100

[1] 99.15533
```

Again, the pathway has a high rate of compliance after intervention despite the relatively poor inspection effectiveness ($\hat{e} = 23\%$) and the small sampling fraction (a 10% sample). As before, the high compliance results from the inherent cleanliness of the pathway; the BIC is 99.13%.

Also, as noted above, PIC has a concrete real-world interpretation, conditional on the assumptions laid out below. An estimated 99.16% of the pathway is compliant with quarantine regulations after intervention.

Example of Planning Future Intervention The pathway manager now has three distinct, but not mutually exclusive, ways to increase the future pathway compliance rate, each of which will be reflected in a different component of the performance indicator.

Summarized briefly, the above results were

PIC: 99.16 %
BIC: 99.13 %
Approach Rate ($\hat{a}_r$): 0.87 %
Inspection Effectiveness ($\hat{e}$): 23.08 %
Pathway Leakage ($\hat{L}$): 844.67 units per year
Inspection Leakage ($\hat{l}$): 66.67 units per year
Volume (v): 100000 units per year.

1. *Increase the inspection rate.* If we, for example, double the inspection rate, then the PIC changes modestly as follows:

   PIC: 99.18 %
   BIC: 99.13 %
   Approach Rate ($\hat{a}_r$): 0.87 %
   Inspection Effectiveness ($\hat{e}$): 23.08 %
   Pathway Leakage ($\hat{L}$): 824.67 units per year
   Inspection Leakage ($\hat{l}$): 133.33 units per year
   Volume (v): 100000 units per year.

2. *Reduce the approach rate.* If we, for example, halve the approach rate via a well-targeted public relations campaign, then the PIC changes as follows:

   PIC: 99.58 %
   BIC: 99.57 %
   Approach Rate ($\hat{a}_r$): 0.43 %
   Inspection Effectiveness ($\hat{e}$): 23.08 %
   Pathway Leakage ($\hat{L}$): 422.33 units per year
   Inspection Leakage ($\hat{l}$): 33.33 units per year
   Volume (v): 100000 units per year.

3. *Increase inspection effectiveness.* If we, for example, approximately double the inspection-level effectiveness by further training, incentive schemes, improving inspection environment, etc, then

   PIC: 99.17 %
   BIC: 99.13 %
   Approach Rate ($\hat{a}_r$): 0.87 %
   Inspection Effectiveness ($\hat{e}$): 46.09 %
   Pathway Leakage ($\hat{L}$): 826.43 units per year
   Inspection Leakage ($\hat{l}$): 46.78 units per year

Volume (v): 100000 units per year.

These calculations clearly show that halving the approach rate is the most useful activity of the three. Of course, this example doesn't take account of the relative costs or feasibility of these exercises. It would be straightforward and useful to perform a benefit-cost analysis based on some educated guesses about the relevant quantities.

5

Discussion

5.1 Consequences

The proposed measure does not take into account the amount of non-compliance detected. A unit is considered either compliant or not compliant. Further, the measure does not take into account the relative consequences of the different types of non-compliance. The PIC, and the statistics that are associated with it, treat all kinds of non-compliance as being equal. It is the responsibility of the user to firstly define non-compliance for the purposes of counting the results, and second, to interpret the PIC result in the light of the definition of non-compliance. This section discusses some of the issues that arise from these points.

Some classifications will necessarily be subjective and may even seem arbitrary. This concern can be reduced by the issuing of clear work instructions for inspectors, and the provision of training. However, at some level it is inevitable that personal judgment will be needed. In some ways this need may seem to be a disadvantage, as it opens the inspectorate to possible accusations of subjectivity and bias. On the other hand, the opportunity for personal judgment allows for the fact that there is no way that work instructions can cover every contingency of non-compliance. Trusting inspectors to make good decisions, and supporting that process by making the decisions as easy as possible, seems to be the best approach.

Also, the classification relies on a simplification that, so far as the performance measure is concerned, either all kinds of non-compliance are equal, or at least that the variation in the rate and the implications of the consequences non-compliance averages out in a meaningful way along the pathway. This simplification is both a very useful and a highly questionable one, and opens the inspectorate to questions like: do we really care as much about a kilogram sample of coffee as a sea container of coffee beans? It seems that the only way to answer this question is to say: yes and no. If the contaminations are on different pathways, such as in this example, then different levels could be required as cutoffs. Pathways in which the non-compliance was considered to be very important could have higher compliance cutoffs than pathways in which the non-compliance is considered less important. For example, they could reflect the assessments of consequences in IRAs (for instance), and be discounted for estimated probabilities of establishment and spread.

An alternative is to follow the approach presently used in Passenger and Mail, which is to distinguish between *risk* and *high-risk* contamination. A PIC could then be computed for both types of contamination, with a concomitant increase in the complexity of the reported measure. PIC can be calculated for any subset of pathways, although it would take longer to obtain sufficient leakage survey data to give an accurate estimate. Further distinctions could be made as necessary.

That said, the identification of non-compliance, and the response to non-compliance, will vary from pathway to pathway. That variation has always been present, but the adoption of a performance measure that responds directly to such considerations will make comparisons more transparent.

5.2 Data Requirements

A major advantage of the PIC as a performance indicator for AQIS is that an estimate of PIC can be made using inspection data that are readily available in most inspection pathways. The data requirements are about as modest as they can be, given that AQIS accepts that inspections may be imperfect and that inspection leakage should be estimated and accounted for.

The original brief for this project was to develop one or more performance indicators that focus on the core operations and how those can be best measured, without regard to existing data holdings or the current ways in which the inspectorate carries out its responsibilities. We think that the PIC and its components are the best possible index, because of its direct relevance to the concrete operations that the inspectorate undertakes. It is serendipitous that the PIC can be computed with existing data for many AQIS pathways.

5.3 Reflecting Uncertainty

The PIC will most often be calculated using information that has come from samples, and hence will be subject to statistical uncertainty. We can therefore calculate not just the estimate of the PIC but also the confidence limits for the estimate.

It is useful to have confidence limits for a PIC so that a person can see how accurate the estimate is, and possibly suggest that a larger leakage survey needs to be done to obtain sufficiently accurate information more quickly.

Example of calculation of a confidence interval for PIC and the components of interest can be found in Appendix B.

If the preferred use of the PIC is to report trends in compliance, then it may also be useful to compute a confidence interval for the change of PIC (δ-PIC) in order to provide guidance as to the statistical quality of the information about the trend. Such an approach would also be useful for selecting a timeframe for review. The utility of this approach will depend on specific characteristics of the pathway, and is the subject of further ACERA work (Project 1101B).

5.4 Measures of Efficiency

The proposed performance indicator focuses on the level of compliance of the pathway after intervention. It takes no direct account of the efficiency of that intervention, that is, the amount of effort that has been undertaken to achieve the compliance level. The efficiency of an inspection operation is an element of implementation, and as such is outside the scope of the current report. However, the topic is sufficiently important to justify some comments.

One efficiency-related measure that AQIS might adopt corresponds to the "specificity" of a serological test. When a blood test for antibodies to an animal disease is performed, the test should be "sensitive" (i.e., pick up as many infected animals as possible) and "specific" (i.e., not setting off many false alarms). The best tests have both sensitivity and specificity close to 100%.

The sensitivity of a test is exactly the same as AQIS "effectiveness": the proportion of infected animals that are detected. AQIS tends to worry less about "specificity": one minus the false positive rate. If the x-ray suggests a problem, AQIS looks at the unit further, by an almost perfect method. Such an approach only becomes a problem if the definitive inspection is time-consuming and there are too many false alarms.

Usually sensitivity and specificity are related: increasing the sensitivity will usually decrease the specificity (and vice versa). The profile list in air cargo is a good example of this: if we remove a profile we might remove a large number of false alarms, but introduce some new 'fail-to-detect'; if we add a profile we increase the effectiveness but might also increase the number of false alarms.

Suppose i units are x-rayed (mail units, passenger bags). Of the suspicious units, c are false alarms and r are BRM. The leakage survey suggests the effectiveness of x-rays is e so that there would be about r/e units with BRM approaching. Looking at how c/i changes is a quick way to monitor false alarms and avoids quirks of the leakage survey.

A more precise way to measure the "specificity" of x-rays would be

$$s = \frac{i - r/e - c}{i - r/e} \tag{5.1}$$

i.e., the proportion of the estimated number of non-BRM units that pass through the x-ray machine without being inspected further.

Where profiling is involved in intervention, it might also be useful to report the benefit of profiling as a ratio of probability of intercepting non-compliant units, as done for example by US Customs.

5.5 Leakage Surveys

As noted earlier, for the purposes of this report, we have assumed that the leakage survey is an intrinsic, continuous, and routine element of the intervention. This assumption has two parts: that the leakage survey is a routine part of the intervention, and that it is continuous, as for example in the Passengers and Mail pathways.

The first part of the assumption is not straightforward, and not universally accepted, even among the authors. Evaluating the performance of intervention is an integral part of intervention and so the leakage survey is indeed an integral part of intervention. Nonetheless, the leakage survey may be thought to only exist to assess the quality of the intervention. Hence any units found in the leakage survey would otherwise have been leaked and cannot be considered as having been "detected" by the intervention.

The difference between the two positions is that in the first case, the non-compliant units that are intercepted in the leakage survey are subtracted from the estimated intervention leakage, whereas in the second case, those non-compliant units are still themselves counted as leakage. In an operational setting the difference will be minor, but the difference in the mathematical model needs to be explained. In the end we retained the assumption that the leakage survey is both routine and continuous, and pointed out where in the development of the model it would make a difference, and provided some insight as to the expected effect of the two different positions.

Also, we have assumed that the leakage survey is being done regularly, though not necessarily every day. The key thing is that when the leakage survey is being done, each unit inspected has (more or less) the same chance of being sampled. We have also assumed that conditions are more or less the same over the period of analysis. However, there will be times when these assumptions are not met. We might only need to do seasonal surveys or we may wish to target our survey. Each will have different implications for estimating the PIC, and a detailed discussion is beyond the scope of this report. We recommend that any changes that are proposed for the leakage survey be considered in the light of their expected impact on the estimate of the PIC, and its confidence interval. That is, the PIC itself can and should be used to assess the likely impact of a change in the leakage survey upon the quality of information that is available to the pathway manager.

For example, if the leakage survey in the pathway described in Section 4.4.1 were to be switched to a seasonal approach, then the pathway manager might ask an analyst to verify whether equation (4.2), or its equivalent in Section B.1, would still provide a suitable estimate of the leakage. If so, then no change would be required. If not, then it would need to be altered. No other aspect of the calculation of the PIC would need to be changed.

5.6 Different Measures

No one measure is ever going to give the complete picture. For example, the quantity of leakage L is important for influencing the frequency of post-border outbreaks (compliance or efficacy can stay the same or even increase, but if L goes up then risk goes up).

The value of L can determine compliance and effectiveness targets — if L is high, it might be worth spending more to get a 0.1% increase in compliance — if L is low, then the same percentage increase is less cost-effective.

Some other measures related to PIC are statistically identical but will yield different interpretations. For example, instead of the post-intervention compliance rate, AQIS could use the post-intervention non-compliance rate, as used in recent ACERA reports (See, e.g., Robinson et al. 2010b, c, d, and e). The post-intervention non-compliance rate would be computed as 1 – PIC. Another alternative is the estimated post-intervention leakage count ($\hat{L}$).

As noted above, performance indicators should reflect the goals of the organization that reports them and AQIS's goal is to minimize the risks of exotic pests and diseases entering the country. Choosing between these different measures reflects sociopolitical risk rather than actuarial risk (see, e.g. Haines et al., 2008; Haines, 2011).

Communication of risks usually involves conveying estimates of the probability of various events to diverse stakeholder groups. An analyst can choose the present information as an incidence rate, a relative risk or a natural frequency (Table 5.1).

Regulators have long understood that people have very different understandings of the concept of probability and frequently misunderstand probability information (Budescu et al., 1988; Osman and Shanks, 2005). Even professionals within the same domain will interpret probabilities in a different or inconsistent fashion (Caponecchia, 2009). This is especially true of medicine and epidemiology (Feldman-Stewart et al., 2000; Viscusi and Aldy, 2003), whose risk profile has a great deal in common with quarantine. This can be a significant strain on time and resources and delay the progress of risk management.

Table 5.1: Alternative formulations of a comliance measure for a scenario in which there is a risk of 'failure' before quarantine measures are imposed (p_1) and a reduced risk following the application of measures (p_2).

Representation	Relevant Quantities	Message
Rate (compliance)	$1 - p_1 = 0.980$; $1 - p_2 = 0.989$	The probability of compliance has increased by 0.9%
Rate (non-compliance)	$p_1 = 0.020$; $p_2 = 0.011$	The probability of non-compliance has fallen by 0.9 percentage points or 45%
Relative risk	$p_2/p_1 = 0.55$	The risk has declined by 55%
Frequency (compliance)	980 in 1000; 989 in 1000	Before, 980 in 1000 consignments passed; after, 989 in 1000 pass.
Frequency (non-compliance)	20 in 1000; 11 in 1000	Before, 20 in 1000 consignments failed; after, 11 in 1000 fail.

Despite this, there are few guidelines on how to convey probability information in such a way that the substance of a message is understood in the way that the communicator intends, by most, if not all, stakeholders.

Epidemiologists have had long and vigorous debates about the right way to present health probability information. There is no single 'right' answer. While the formats in Table 5.1 do not provide protection against misinterpretation, Gigerenzer et al. (2005) suggested that many such problems may be reduced if care is taken to communicate the 'reference class' for a probability statement. The reference class is usually described in additional information supplied along with a probability statement, and indicates the situations to which the probability applies, the class of events or conditions from which the probability can be thought of as a kind of random draw.

Framing refers to the way in which a message is communicated, leading people to draw different conclusions from the same information. Availability bias is the propensity for people's judgments to be influenced more heavily by experiences that most easily come to mind. Motivational bias arises when people consider what they have to gain or lose personally from an unwanted event. If a reference class is not supplied, stakeholders will infer their own reference class, often incorrectly. Framing information, availability and motivational bias will heavily influence them (Gigerenzer, 2002).

As Table 5.1 illustrates, some measures related to PIC are statistically identical but will yield different interpretations. For example, instead of the post-intervention compliance rate, AQIS could use the post-intervention non-compliance rate, as used in recent ACERA reports (See, e.g., Robinson et al., 2010b, c, d, e). The post-intervention non-compliance rate would be computed as 1 − PIC.

A disadvantage of PIC in some circumstances may be that the relative increase in PIC for process improvement can seem poor. Referring to the examples in Section 4.4.2, recall that the outcome of doubling the inspection rate was a modest change of 99.18% to 99.19%. The PIC increases by just 0.01%, which seems like a minor change operationally, and would require a considerable amount of data to demonstrate statistically. If post-intervention non-compliance were used as a measure, then the change would be from 0.82% to 0.81%, which is an absolute decrease of 0.01%, but a relative decrease of just over 1%. This is still small, but gives a greater sense of achievement in the context of the pathway.

Another alternative is the estimated post-intervention leakage count ($\hat{L}$). If the estimated leakage count were used to evaluate the expected benefit, the change would be from 823 to 815 estimated contaminated units post-intervention, equating to an estimated 8 extra interceptions. This measure provides a different perspective on the benefit of the extra inspection, especially when considered with the cost of that extra intervention. Fortunately, all the measures noted above are commensurate, meaning that they are internally consistent and any can be reported without fear that the statistics themselves will disagree.

We recommend that for any given context and circumstance, the communicator should consider the intent of the communication and the context and motivational biases of the target audience, and then choose an appropriate frequency format and reference class. In most circumstances, this will involve describing the reference class carefully, and communicating probability information as natural frequencies. A range of other visual and numerical tools may assist a communicator to shape a message so that it reaches the target audience in an effective form (Caponecchia, 2009).

5.7 Hierarchical Scale of Reporting

When designing, carrying out, and assessing interventions, it is important to think about the scale of the harms and the solutions (Sparrow, 2008). In the context of the current report, we should ask: when prescribing a performance measure, what is/are the hierarchical level(s) at which it should be reported, to whom should it be reported, and how frequently?

The Air Cargo program provides an artificial example of this problem: should PIC be reported separately for the reportable documents pathway and the universal loading devices pathway? Here, the answer is yes. Should PIC be reported for each of the regions or nationally, or both? Here, the answer might be: both. Less clear cut is whether we should calculate a PIC for each carrier — we should probably not unless there are operational differences between carriers, because monitoring the detection rate as a de-facto measure for approach rate provides adequate information about the risk associated with each carrier. Also, the smaller sample size that will inevitably be associated with carriers (as opposed to the pathway itself) will result in inflated uncertainty, perhaps unreasonably so. Recall, for example, that 600 inspections must all detect no non-compliance in order for the upper 95% confidence limit of the estimated non-compliance rate to be below 0.5%.

In contrast, in the Plants program, should PIC be reported for each of the approximately 50,000 commodities that are monitored? Here, the answer might be *yes* if the level of intervention being done is determined by the PIC, or *no* based on the large number of pathways that would force the managing program to handle, especially as the uncertainty that propagates to the PIC from small leakage samples would make confidence intervals for the performance indicator (for example) very wide. Aggregating the results to higher hierarchical levels where sensible, even if just for estimating leakage, will reduce uncertainty considerably. For example, aggregating the commodities into like groups, somehow, and considering compliance for those groups, might work. These comments point out the arbitrariness of the definition of a pathway.

There is no simple answer to this question, and possibly, the question can be answered operationally. Operational considerations will usually determine the reporting level and frequency. Program managers will require more frequent and more detailed performance indicators of narrower pathways than will the executive.

The frequency of reporting will also depend on how finely we are breaking down a pathway because the accuracy of the estimated statistics depends on the size of the leakage surveys. For example, the leakage surveys might be large enough to make monthly national statistics meaningful, but regional statistics would require leakage survey data across three months. Then aggregation across longer time periods provides a painless way to obtain more information on smaller sub-pathways.

5.8 A Word in Favour of Detection Rates

We noted earlier that the ANAO criticized the use of only the number of detections or seizures as a performance indicator, because it failed to take account of the amount of effort undertaken by the inspectorate, and the effectiveness of the effort. However, under the assumption that the amount of effort and the effectiveness of effort are reasonably stable over medium time periods, the detection or seizure rate is an extremely easy way to monitor performance over a short time period.

There is one problem with using effectiveness as a management tool over a short period of time: statistical variation in the (usually small) leakage survey can lead to quite variable estimates of effectiveness. This variation is reduced as the information from a number of leakage surveys is accumulated over time. False alarms about poor performance are of particular concern to managers and staff alike.

Because the number of detections is based on all the units inspected, changes will reflect actual changes rather than being a quirk of the sample results. Since the number of units inspected is the greatest factor in determining the number of detections made, it is best to use detection rate (rather than number of detections). It is fair to conclude that, if the detection rate has stayed the same despite a temporary drop in effectiveness as measured by a leakage survey, there is not likely to be a problem.

On a more clerical note, daily monitoring of detection rates and other throughput statistics can alert managers to data entry or collection failures for which there is still time to take remedial action.

Therefore, we note that regular monitoring of detection rates is a very useful short-term adjunct to the PIC. Notwithstanding the comments above, monitoring effectiveness is still a valuable tool for managers to evaluate intervention performance.

5.9 Comparing PIC with a Target

One way to interpret the PIC is relative to a target. In choosing a target for PIC it seems politically expedient to decompose the challenge into two questions: first, will the community require 100% compliance, and second, if not, then to what level should compliance be set? This decomposition distinguishes between the two distinct problems of: will we allow any risk at all, and if so, then how much?

Setting a PIC target of less than 100% allows for the unpleasant inevitability that some contamination will enter the country. However, as noted by Beale et al. (2008, p. xvi), “Zero risk is unattainable and undesirable.” Given that the risk can’t be zero, or equivalently, the compliance can't be 100%, the community needs to set a lesser target for compliance.

This dilemma is true of any performance target for inspections. There is no simple answer to this question, and possibly, the question will have to be answered operationally. As outlined in many reviews of AQIS, the response to a threat should be commensurate with the consequences. Hence, the target may be different for different programs. The target should reflect the volume of the pathway, because the PIC is a rate of compliance.

5.10 Computation Notes

When calculating a PIC (and some other measures like approach rate and leakage) it is best to keep them as fractions within the spreadsheet, and use the formatting capabilities to display them as percentages. This strategy reduces conversion error.

6

Recommendations

We make the following recommendations.

1. That AQIS use the PIC and its components as a performance indicator for the international passengers and mail pathways.
2. That AQIS assess the utility of the PIC and its components as a performance indicator for other pathways, such as air cargo, and plant products (ACERA Project 1101D: Adoption of Meaningful Performance Indicators).

Bibliography

ANAO (2001). Managing for quarantine effectiveness: Department of Agriculture, Fisheries and Forestry, Biosecurity Australia. Audit report no. 47, 2000–01 Performance Audit, Australian National Audit Office, Commonwealth of Australia, Canberra.

Beale, R., Fairbrother, J., Inglis, A., and Trebeck, D. (2008). *One Biosecurity: a Working Partnership*. Commonwealth of Australia.

Budescu, D. V., Weinberg, S., and Wallsten, T. S. (1988). Decisions based on numerically and verbally expressed uncertainties. *Journal of Experimental Psychology: Human Perception and Performance*, 14:281–294.

Caponecchia, C. (2009). Strategies to improve the communication of probability information in risk analyses. *Int. J. Risk Assessment and Management*, 12:380–395.

Feldman-Stewart, D., Kocovski, N., Mcconnell, B. A., Brundage, M. D., and Mackillop, W. J. (2000). Perception of quantitative information for treatment decisions. *Medical Decision Making*, 20:228–238.

Gigerenzer, G. (2002). *Calculated risks: how to know when numbers deceive you*. Simon and Schuster, New York.

Gigerenzer, G., Hertwig, R., van Den Broek, E., Fasolo, B., and Katsikopoulos, K. V. (2005). 'a 30% chance of rain tomorrow': how does the public understand probabilistic weather forecasts? *Risk Analysis*, 25:623–629.

Haines, F. (2011). *The Paradox of Regulation: What Regulation Can Achieve and What It Cannot*. Edward Elgar, Cheltenham.

Haines, F., Sutton, A., and Platania-Phung, C. (2008). It's all about risk, isn't it? science, politics, public opinion and regulatory reform. *Flinders Journal of Law Reform*, 10(3):435–453.

Osman, M. and Shanks, D. R. (2005). Individual differences in causal learning and decision making. *Acta Psychologica*, 120:93–112.

R Development Core Team (2010). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

Robinson, A., Burgman, M., Atkinson, W., Cannon, R., Miller, C., and Immonen, H. (2008). AQIS Import Clearance Data Framework. Technical Report 0804, Australian Centre of Excellence for Risk Analysis.

Robinson, A., Karri, R., Clarke, F., and Scott, M. (2010). AQIS Import Clearance Risk Return: RD Report 3. Technical Report 1001f 3, Australian Centre of Excellence for Risk Analysis.

Sparrow, M. K. (2000). *The Regulatory Craft: Controlling Risks, Solving Problems, and Managing Compliance*. Brookings Institution Press.

Sparrow, M. K. (2008). *The Character of Harms: Operational Challenges in Control*. Cambridge University Press.

Viscusi, W. K. and Aldy, J. E. (2003). The value of a statistical life: A critical review of market estimates throughout the world. *Journal of Risk and Uncertainty*, 27:5–76.

Appendix A

Generalized PIC

A.1 h Sub-Pathways

This development is for the case in which the pathway is divided into multiple sub-pathways. We assume that the approach rate, sampling fraction, and effectiveness may be different in the sub-pathways. We use the same assumptions as are documented in Section 4.4. As before, numerous other assumptions are possible, and a comprehensive coverage is beyond the scope of this document.

We use the same definitions as in Section 4.4, but now we add a subscript to the variable labels: k, representing the sub-pathway number. The inspection leakage is estimated by

$$\hat{l}_k = \frac{i_k \times y_k}{n_k} \tag{A.1}$$

The estimated inspection effectiveness $\hat{e}$ is then the ratio of what was found over the estimated non-compliance among the inspected units.

$$\hat{e}_k = \frac{b_k}{b_k + \hat{l}_k} \tag{A.2}$$

Now the estimated approach counts $\hat{a}_k$ for the sub-pathways are

$$\hat{a}_k = (b_k + \hat{l}_k)\frac{v_k}{i_k} \tag{A.3}$$

Then

$$\hat{L}_k = \hat{a}_k - b_k - y_k \tag{A.4}$$

So,

$$\hat{L} = \sum_h \hat{L}_k \tag{A.5}$$

and

$$PIC = \frac{v - \hat{L}}{v} \tag{A.6}$$

where

$$v = \sum_h \hat{v}_k \tag{A.7}$$

Again, in terms of direct measurements, we can compute

$$\hat{L}_k = b_k \times \left(\frac{v_k}{i_k} - 1\right) + y_k \times \left(\frac{v_k}{n_k} - 1\right) \tag{A.8}$$

or, if the leakage survey is not an integral part of the intervention, use

$$\hat{L}_k = b_k \times \left(\frac{v_k}{i_k} - 1\right) + y_k \times \left(\frac{v_k}{n_k}\right) \tag{A.9}$$

Finally, we may wish to know

$$\hat{a} = \sum_h \hat{a}_k \text{ and } \hat{l} = \sum_h \hat{l}_k \tag{A.10}$$

However, the $\hat{e}_k$ are not additive in the same way. The overall effectiveness can be computed by

$$\hat{e} = \frac{\sum_h b_k}{\hat{a}} \tag{A.11}$$

A.2 Different leakage survey scenarios

This section estimates the approach rate, the effectiveness of inspection, and the PIC for two common variants of how the leakage survey is done. Other variants that can be analysed using the methods of this section. The same notation is used as in the rest of the report.

A.2.1 Partial Rectification.

The first difference to the other parts of this report occurs if not all the inspected units re-enter the pathway. They may have been destroyed or undergone such a thorough inspection or treatment that it is not thought necessary for them to be included in the leakage survey. For example, a mail unit, deemed suspicious on the basis of x-raying, might be opened for a full inspection (equivalent to a leakage survey) and the unit not returned to the normal exit stream. We shall denote the number of inspected units that do not re-enter the exit stream by D.

As before, the seizure rate B/I provides an estimate of $a_r \times e$ (i.e. approach rate times effectiveness of inspection). If all units with BRM are rectified and returned to the output stream with a chance to be included in the leakage survey then, as Section 4.4 showed, the observed leakage (rate) y/n provides an estimate of $a_r \times (1 - e)$. However, if D units do not return to the exit stream, this proportion of units with BRM at the time of the leakage survey must be increased by a factor of $I/(I - D)$: there are the same number of missed units but in slightly fewer units.

Hence the estimated number of units missed ($\hat{l}$) by the inspection is

$$\hat{l} = (I - D) \times \frac{y}{n} = a_r \times (1 - e) \times I \tag{A.12}$$

We can replace $ar \times e$ by its estimate B/I, shuffle the equation around, and find that

$$\hat{a}_r = (\hat{l} + B)/I \tag{A.13}$$

A.2.2 Leakage Survey Of Almost All Units

Our second common variant is that all but the D "destroyed" units are eligible for the leakage survey, whether inspected on not. Such a situation might apply to air cans in which the leakage inspection is done without any knowledge of the inspection history of the air can.

In such a case, the expected number of units with BRM at the leakage survey is $(v-I)\times a_r$ units having BRM in the $(v-I)$ uninspected units plus $I\times a_r\times(1-e)$ units with BRM in the $(I-D)$ inspected units. Hence the total number of units missed $(\hat{l})$ by the process (note the slight change of definition of $\hat{l}$) is:

$$\hat{l} = (v-D)\times\frac{y}{n} = (I-D)\times a_r\times(1-e)+(v-I)\times a_r \tag{A.14}$$

After replacing $a_r\times e$ by its estimate B/I we can simplify this to obtain

$$\hat{a}_r = \frac{\hat{l}+(I-D)\times B/I}{N-D} \tag{A.15}$$

For both these examples we can then calculate

$$\hat{e} = B/I/\hat{a} \tag{A.16}$$

and $PIC = 1-\hat{a}+$ either B/v or $(B+y)/v$ depending on how the leaked items are included in the PIC.

Appendix B

Confidence Intervals for PIC using R

As mentioned earlier, confidence intervals are useful tools to aid in interpreting the precision of estimates of quantities of interest. In general, two kinds of intervals are computed: confidence intervals, which refer to the statistical uncertainty of a process or population parameter, and prediction intervals, which refer to the statistical uncertainty of an observation. Given that our goal is to communicate the uncertainty about inspectorate performance as a process, we advocate that confidence intervals be reported, as opposed to the wider prediction intervals.

Here we demonstrate one approach to the calculation of relevant confidence interval estimates for statistics of interest in the calculation of the PIC, including the PIC itself, for the two cases presented in Section 4.4. Other approaches are possible, and further research into these different approaches is underway within ACERA. We use the free, open-source statistical environment R (R Development Core Team, 2010). All the code to run these examples is available from the first author.

Our general approach will be to use the same equations as above, but replacing random variables for the constant values whenever the true value is known only with uncertainty.

The PIC and all the other statistics relating to the intervention depend on two fundamental parameters: the approach rate of BRM, which we call a_r, and the effectiveness, which we have called e. The number of units with BRM found by inspectors (which we have called b) is a random variable, to be specific a binomial random variable with a probability of $a_r \times e$ that a single inspection will find BRM. Similarly, the number of units of BRM (y) found in the leakage survey has a binomial distribution but with a probability of finding BRM: $a_r \times (1 - e)$. This probability assumes that units found by inspection are rectified and then might be sampled in the leakage survey. The probability is the more complicated $a_r \times (1 - e)/(1 - a_r \times e)$ if this is not so because the units found are either destroyed or, if rectified, are ignored by the leakage survey. The equations of Section 4.4 manipulate the data so that the a_r and the e are untangled from these two probabilities.

To determine confidence intervals we look at how likely the observed results of b and y would occur for different values of a_r and e. Our confidence limits for a_r and e are those values for which the likelihood of observing b and y isn't too low. Because b has a binomial distribution, we know the relative likelihood of finding b units for each possible value of $a_r \times e$, although we can't separate the a_r and e without more data. We shall use `b_d` to represent the likelihood of $a \times e$, the `_d` part of the name being to remind us that it is a likelihood and the "b" part to remind us what we have the likelihood of. If we inspect i units and find b then the likelihood `b_d` has the shape of a beta distribution: `b_d` $\sim \text{beta}(b + 1, i - b + 1) \times i$. Similarly the likelihood of y units being found in the n

units in the leakage survey is `y_d` beta$(y+1, n-y+1) \times n$. There are several ways that the two likelihood functions `b_d` and `y_d` can be manipulated to make inferences about the values of a_r, e and PIC. For this Appendix we shall use the R programming language which has packages that enable the manipulation of likelihood functions.

Hence we will create two new random variables, `b_d` and `y_d`, which represent the non-conformities detected in the usual inspection and the leakage survey respectively, but also carry information about the approximate statistical uncertainty of the information.

The R programming language provides packages that enable the manipulation of probability densities as variables. These packages are called `distr` and `distrEx`. The packages also provide functions `E(X)` and `q(X)` to calculate the mean and quartiles of the random variable X respectively. In order to use these we must begin by invoking the packages as follows.

```
> library(distr)
> library(distrEx)
```

B.1 Completely Inspected Pathway

We begin with the pathway for which all the inspection units are inspected, and for which a leakage survey is available. See page 26 for a brief explanation of the R output. The example, fabricated data previously used were

```
> v = 10000
> b = 20
> n = 300
> y = 2
```

We create the random variables by

```
> b_d = Beta(shape1 = b + 1, shape2 = v - b + 1) * v
> y_d = Beta(shape1 = y + 1, shape2 = n - y + 1) * n
```

Then the estimated inspection leakage was previously

```
> (l_hat = v * (y/n))
```

```
[1] 66.66667
```

however, because `Y` is really a binomial random variable (not a constant), `l_hat` is also a random variable. Therefore we write

```
> l_hat_d = v * (y_d/n)
```

Now we ask R to report important characteristics of the random variable: specifically the mean, median, and the 95% confidence interval, as follows.

```
> E(l_hat_d)
```

```
[1] 99.33775
```

```
> q(l_hat_d)(c(0.025, 0.5, 0.975))
```

```
[1]  20.60124  88.73981 237.94872
```

Then the estimated inspection effectiveness (reported as a percentage) was

```
> (e_hat = b/(b + l_hat)) * 100
```

```
[1] 23.07692
```

but again we replace `l_hat` by `l_hat_d` so that the effectiveness is a random variable: `e_hat_d`. We also need to re-express the equation so that each random variable appears only once. Note that if

$$e = \frac{b}{b + l}$$

then it is also true that

$$e = \frac{1}{\frac{l}{b} + 1} \tag{B.1}$$

however, in (B.1) the random variables each appear only once. This step is necessary because the instances of b created by R would be independent, and would inflate the uncertainty of the resulting random variable.

We obtain important information about `E_hat_d` as follows.

```
> e_hat_d = 1/(l_hat_d/b_d + 1)
> E(e_hat_d) * 100
```

```
[1] 21.58016
```

```
> q(e_hat_d)(c(0.025, 0.5, 0.975)) * 100
```

```
[1]  7.223852 18.887898 51.265435
```

Note that the mean and the original estimate are quite different. This is due at least in part to Jensen's Inequality, which states that the mean of a curved function of data is not the same as the curved function of the mean of the data. Hence, the mean of `e_hat_d` is probably a better estimate than `e_hat` is.

The estimated approach count was

```
> (a_hat = b + l_hat)
```

```
[1] 86.66667
```

but, using a similar argument, is now

```
> a_hat_d = b_d + l_hat_d
> E(a_hat_d)
```

```
[1] 120.3332
```

```
> q(a_hat_d)(c(0.025, 0.5, 0.975))
```

```
[1]  40.67389 109.83846 259.25357
```

The estimated pathway leakage *count* was

```
> (L_hat = y * (v/n - 1))
```

```
[1] 64.66667
```

and now is

```
> L_hat_d = y_d * (v/n - 1)
> E(L_hat_d)
```

```
[1] 96.35762
```

```
> q(L_hat_d)(c(0.025, 0.5, 0.975))
```

```
[1]  19.98321  86.07762 230.81026
```

or, if the leakage survey is not part of the intervention,

```
> L_hat_d2 = y_d * (v/n)
> E(L_hat_d2)
```

```
[1] 99.33775
```

```
> q(L_hat_d2)(c(0.025, 0.5, 0.975))
```

```
[1]  20.60124  88.73981 237.94872
```

Finally the PIC, which was

```
> (PIC = (v - L_hat)/v)
```

```
[1] 0.9935333
```

is now

```
> PIC_d = (v - L_hat_d)/v
> E(PIC_d)
```

```
[1] 0.9903642
```

with 95% confidence interval

```
> q(PIC_d)(c(0.025, 0.5, 0.975))
```

```
[1] 0.9769190 0.9913922 0.9980017
```

B.2 Sampled Pathway

We provide a similar example for the pathway for which units are only sampled. We will keep the commentary brief, relying on the reader to refer to the previous section.

```
> v = 1e+05
> i = 10000
> b = 20
> n = 300
> y = 2
```

```
> b_d = Beta(shape1 = b + 1, shape2 = i - b + 1) * i
> y_d = Beta(shape1 = y + 1, shape2 = n - y + 1) * n
```

The estimated inspection leakage is:

```
> (l_hat = i * (y/n))
```

```
[1] 66.66667
```

```
> l_hat_d = i * (y_d/n)
> E(l_hat_d)
```

```
[1] 99.33775
```

```
> q(l_hat_d)(c(0.025, 0.5, 0.975))
```

```
[1]  20.60124  88.73981 237.94872
```

The estimated inspection effectiveness (reported as a percentage) is

```
> (e_hat = b/(b + l_hat)) * 100
```

```
[1] 23.07692
```

```
> e_hat_d = 1/(l_hat_d/b_d + 1)
> E(e_hat_d) * 100
```

```
[1] 21.58016
```

```
> q(e_hat_d)(c(0.025, 0.5, 0.975)) * 100
```

```
[1]  7.223852 18.887898 51.265435
```

The estimated approach count is

```
> (a_hat = (b + l_hat) * (v/i))
```

```
[1] 866.6667
```

```
> a_hat_d = (b_d + l_hat_d) * (v/i)
> E(a_hat_d)
```

```
[1] 1203.332
```

```
> q(a_hat_d)(c(0.025, 0.5, 0.975))
```

```
[1]  406.7389 1098.3846 2592.5357
```

The estimated pathway leakage *count* is (in terms of the original random variables, which is better for the simulation),

```
> (L_hat = b * (v/i - 1) + y * (v/n - 1))
```

```
[1] 844.6667
```

```
> L_hat_d = b_d * (v/i - 1) + y_d * (v/n - 1)
> E(L_hat_d)
```

```
[1] 1179.348
```

```
> q(L_hat_d)(c(0.025, 0.5, 0.975))
```

```
[1]  386.2762 1074.5793 2563.9154
```

or, if the leakage survey is not part of the intervention,

```
> (L_hat2 = b * (v/i - 1) + y * (v/n))
```

```
[1] 846.6667
```

```
> L_hat_d2 = b_d * (v/i - 1) + y_d * (v/n)
> E(L_hat_d2)
```

```
[1] 1182.33
```

```
> q(L_hat_d2)(c(0.025, 0.5, 0.975))
```

```
[1]  386.9169 1077.2417 2571.0598
```

Finally, the PIC is

```
> (PIC = (v - L_hat)/v)
```

```
[1] 0.9915533
```

```
> PIC_d = (v - L_hat_d)/v
> E(PIC_d)
```

```
[1] 0.9882065
```

```
> q(PIC_d)(c(0.025, 0.5, 0.975))
```

```
[1] 0.9743608 0.9892542 0.9961372
```

B.3 Multiple Sub-Pathways

We conclude by developing an R function[1] that automates this output for multiple sub-pathways, and demonstrating it on some example data.

Here we report L as the pathway-level number of units leaked from inspection and a_r as the approach rate relative to v. We emphasize that these statistics are purely for the purposes of providing an example, and bear no necessary relation to any inspection data.

```
> (pathway <- data.frame(vk = c(10000, 1e+05), ik = c(10000,
+     10000), bk = c(20, 20), nk = c(300, 300), yk = c(2,
+     2)))

     vk    ik bk  nk yk
1 1e+04 10000 20 300  2
2 1e+05 10000 20 300  2

> rownames(pathway) <- paste("Pathway ", 1:2, sep = "")

> (example <- pic(pathway))

                   l_hat     l_50    l_lcl    l_ucl
Pathway 1       99.33775 88.73981 20.60124 237.9487
Pathway 2       99.33775 88.73981 20.60124 237.9487
Entire Pathway        NA       NA       NA       NA

                    e_hat     e_50      e_lcl     e_ucl
Pathway 1       0.2158016 0.188879 0.07223852 0.5126543
Pathway 2       0.2158016 0.188879 0.07223852 0.5126543
Entire Pathway         NA       NA         NA        NA

                    ar_hat      ar_50      ar_lcl     ar_ucl
Pathway 1       0.01203332 0.01098385 0.004067389 0.02592536
Pathway 2       0.01203332 0.01098385 0.004067389 0.02592536
Entire Pathway 0.01203228 0.01109724 0.004639126 0.02471736

                     L_hat       L_50     L_lcl     L_ucl
Pathway 1         96.35762   86.07762  19.98321  230.8103
Pathway 2       1179.34835 1074.57928 386.27615 2563.9154
Entire Pathway 1275.52430 1172.56712 463.93962 2666.0060

                  PIC_hat    PIC_50   PIC_lcl   PIC_ucl
Pathway 1       0.9903642 0.9913922 0.9769190 0.9980017
Pathway 2       0.9882065 0.9892542 0.9743608 0.9961372
Entire Pathway 0.9884043 0.9893403 0.9757636 0.9957824
```

The output includes the mean (estimate, labeled with "`hat`"), median, and approximate upper and lower limits of a 95% confidence interval for the following quantities: the inspection leakage (`l`), the inspection effectiveness (`e`), the approach rate of non-compliance (a_r, `ar` in the output), the pathway-level leakage (`L`), and the PIC (`PIC`).

---

[1] Available from first author

B.4 Allocation Example

Now we assess the effect upon the distribution of the PIC of different investments of inspection resources. We compare two variants of two scenarios, specified as follows. All scenarios have the same volume of 10000 units.

1. 10% sample, 1% failure rate, 1% leakage survey rate, and 1% contamination rate in the leakage survey. The two variants are:
   (a) Add 100 units to the leakage survey, and
   (b) Add 100 units to the inspection.
2. 20% sample, 1% failure rate, 25% leakage survey rate, and 0.2% contamination rate in the leakage survey. The two variants are:
   (a) Add 500 units to the leakage survey, and
   (b) Add 500 units to the inspection.

```
> (test <- data.frame(vk = c(10000, 10000, 10000, 10000,
+     10000, 10000), ik = c(1000, 1000, 1100, 2000, 2000,
+     2500), bk = c(10, 10, 11, 20, 20, 25), nk = c(100,
+     200, 100, 500, 1000, 500), yk = c(1, 2, 1, 1, 2,
+     1)))
> rownames(test) <- c("Baseline 1", "M Leak 1", "M Insp 1",
+     "Baseline 2", "M Leak 2", "M Insp 2")
> s1 <- s2 <- 1
> out <- pic(test)
```

Our results are obtained using the prior constants of $s_1 = s_2 = 1$. Scenario 1 is:

```
> out[1:3, c("PIC_hat", "PIC_50", "PIC_lcl", "PIC_ucl")]

              PIC_hat    PIC_50   PIC_lcl   PIC_ucl
Baseline 1 0.9707080 0.9735592 0.9363245 0.9891867
M Leak 1   0.9755656 0.9769350 0.9546259 0.9887836
M Insp 1   0.9708967 0.9737655 0.9365622 0.9892290
```

and Scenario 2 is:

```
> out[4:6, c("PIC_hat", "PIC_50", "PIC_lcl", "PIC_ucl")]

              PIC_hat    PIC_50   PIC_lcl   PIC_ucl
Baseline 2 0.9878233 0.9882468 0.9802841 0.9929260
M Leak 2   0.9889134 0.9891072 0.9836677 0.9930571
M Insp 2   0.9884216 0.9888691 0.9811352 0.9931429
```

These results show that, for these pathways, under the assumptions of the PIC, dedicating the extra units to the leakage survey resulted in a larger increase in the lower limit of the pathway compliance compared with adding the extra units to the standard intervention. The results were reversed for the upper limit of the interval estimate for the PIC. Hence, if the mean, the median, or a lower interval limit of the estimated PIC distribution were used to guide resource allocation, the leakage survey would be a better investment in this case than the usual intervention.

Appendix C

Glossary

C.1 Important Acronyms

ANAO	Australian National Audit Office
APHIS	USDA Animal and Plant Health Inspection Service
BIC	Before Intervention Compliance
BRM	Biosecurity Risk Material
MAF	New Zealand Ministry of Agriculture and Forestry
PIC	Post-Intervention Compliance

Appendix D

Change Log

2 December 2011 Final version with edits from all contributors.

14 September 2011 Updates to reflect SAC and DAFF technical and policy review comments.

4 July 2011 Minor edits and spelling corrections.

28 June 2011 Extraction of Bayesian explanations.

3 May 2011 Respond to comments from Tina Hutchison.

13 April 2011 Review of equations and text.

2 April 2011 Substantial rewriting of PIC derivation thanks to helpful comments from Carolyn Whyte.

6 February 2011 Substantial rewriting due to comments from Rob Cannon (now included as co-author) and Mark Burgman.

21 January 2011 Initial version.