Report Cover Page

ACERA Project		
0804		
Title		
AQIS Import Clearance Review		
Author(s) / Address (es)		
A. Robinson, Department of Mathematics and Statistics, University of Melbourne, M. Burgman, ACERA, University of Melbourne, W. Atkinson, AQIS, R. Cannon, AQIS, C. Miller, AQIS, H. Immonen, AQIS.		
Material Type and Status (Internal draft, Final Technical or Project report, Manuscript, Manual, Software)		
Final Project report		
Summary		
This report outlines a system that will assist AQIS managers to allocate effort among inspection activities. It provides a means of guaranteeing a specified degree of reliability in detection that takes into account the underlying risk, and that also provides information and a degree of deterrence for less risky pathways. The system is provided within a wider framework that will assist managers to feed back improved knowledge on risk to those who set policy, as information accumulates over time.		
ACERA Use only	Received By:	Date:
	ACERA / AMSI SAC Approval:	Date:
	DAFF Endorsement: ( ) Yes ( ) No	Date:

AQIS Import Clearance Data Framework
0804 Final Report

Andrew Robinson, University of Melbourne
Mark Burgman, ACERA
Wayne Atkinson, AQIS
Rob Cannon, AQIS
Chris Miller, AQIS
Hendrik Immonen, AQIS

22 December 2008

Acknowledgements

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 valuable advice and input: Adam Bennet, Elizabeth Bragg, Simon Callow, Damien Carthew, Neil Charles, Andrew Christie, Ferne Clark, Lyell Copeland, David Cox, Tom Delmont, Mandy Dines, Dennis Foden, Frank Fox, Michael Franklin, Dr. Nora Galway, Danny Genn, Ian Hanson, Catherine Hill, Greg Hood, Nicole Horner, Ian Jeffries, Robert Langlands, Melessa Lawton, Professor Chris Lloyd, Marissa McBride, Greg Messiter, Dennis Monahan, Teresa Morahan, David Pisk, Boris Popovic, Adam Pyman, Chez Richards, John Sagona, Norm Scott, Harold Smithard, Moshe Sniedovich, Jackie South, Glen Stewart, John Stickens, Professor Colin Thompson, for almost all of Appendix B, Trish Thornhill, Ian Towers, Roger Turner, Meredith Walton, Margaret Ward, Al Whitby, and Michelle Wickham.

Disclaimer

Contents

List of Tables

List of Figures

1

Executive Summary

This report delivers

- a review of current risk management strategies, data collection and processes (Chapter 3),
- a risk management framework that provides a platform for the Import Clearance Program to identify priorities for dealing with alternative sources of risk (Chapter 4),
- examples and tests of tools developed as part of this work for viewing trends, and making decisions about resource allocation (Chapter 5),
- recommendations on how this proposed system could be put into operation, including recommendations for information collection and processes for analysis (Chapter 6), and
- simple spreadsheet programs that implement the risk management framework and resource allocation solutions.

The report set out to describe simple answers to fundamental questions about how to allocate resources among different commodities and activities, while maintaining the need for AQIS to acquire data about failures in all potential pathways, and to provide a degree of scrutiny to act as a deterrent and to encourage good operational practice among trading partners. The answers contained in this report are simple, but like all such tools, they need to be employed by people who are aware of their limitations, and the conditions and assumptions under which they operate.

This report explores the concepts and analyses several case studies to illustrate the tools. Work remains to be done to ensure that these tools are implemented under operational conditions to greatest effect. The next stage will be to define a wider class of conditions and contexts for which they will be appropriate, and to set up the feedback mechanisms to ensure outputs contribute to the next phases of decision making and resource allocation. It was considered to be beyond the scope of this report to recommend detailed procedures for AQIS officers.

The specific recommendations are

1. The establishment of an internal risk team.
2. The development of a collection of easily-grasped decision scenarios, which may be based on or related to the case studies described in this report, to allow AQIS officers to formally calibrate the risk-based approaches.
3. A simulation study to compare the properties and performance of the different risk-estimation approaches that have been canvassed in this report, including data mining.
4. That a Multi-Criteria Decision Analysis implementation program include tests of alternative value trees, weights and aggregation methods, with the ultimate aim of finalizing a tool that could be used in risk minimization and feedback systems.

5. That AQIS IC undertake risk analyses of the kinds demonstrated for aircans (Section 5.1), white rice (Section 5.2), and HVLV (Section 5.4) for other commodities or processes of interest. Examples of such risk analyses are:
   (a) A risk analysis for ECIR.
   (b) A risk analysis for gun handle inspection.
   (c) A risk analysis for QAP audits.
   (d) A risk analysis for fresh produce inspection.
   (e) A risk analysis for the potential utility of profiling for risk management.
6. We recommend that AQIS IC collect and store data about the identity of each step of the import process.
7. That AQIS IC collect and store data on conformities as vigilantly as it does on non-conformities.
8. That AQIS IC develop and implement flexible data collection, storage, and access protocols to enable risk analyses for a wider range of inspection challenges. Examples of such risk analysis opportunities are:
   (a) A risk analysis for Personal Effects (PE) inspection.
   (b) A risk analysis for the quantity of bark on packaging material.
   (c) A risk analysis for seed contamination of new cars.
9. That AQIS IC research, acquire, and deploy data manipulation tools that will enable the extraction, manipulation, and analysis of the data that are stored but inaccessible in AQIS databases such as AIMS.
10. That AQIS should get joint ownership of the incoming data from ICS.

2

Introduction

2.1 Preamble

Quarantine is about controls to maintain Australia's privileged human, animal and plant health status. Australians generally benefit from a natural environment that, compared to other countries, is relatively free of many debilitating pests and diseases of humans, animals and plants. Effective and efficient quarantine controls enhance the quality of life of all Australians by protecting public health, contributing to Australia's comparative advantage in agricultural production, reducing the need to use chemicals to prevent and control pests and diseases, protecting native flora and fauna and promoting Australia as a tourist attraction. There are a number of different pieces of legislation that relate to Australia's quarantine system. The primary source of legislative power with respect to quarantine is the Quarantine Act 1908, which came into force on 1 July 1909. Although it has retained the same name, this Act has been amended many times since introduction. In summary, the Act provides (for the Commonwealth):

- the legal basis for preventing or controlling the entry of people, vessels goods, animals and plants into Australia
- the legal basis for the managing the quarantine risk arising from people, vessels, goods, animals and plants after arrival in Australia
- the powers for the Director of Quarantine and quarantine officers to deal with quarantine matters; and
- for offences and maximum penalties for contraventions of the Act.

The Act also provides the legal authority for the making other subordinate legislation. This subordinate material consists of a number of different legislative instruments, including:

- the three Quarantine Proclamations: the Quarantine Proclamation 1998; the Quarantine (Cocos Islands) Proclamation 2004 and the Quarantine (Christmas Island) Proclamation 2004; and
- the Quarantine Regulations 2000.

Briefly, the Proclamations and Regulations provide additional information and context around the legislative powers as described in the Act.

2.2 Glossary

The key terms used in this report are defined as follows:

Effectiveness : Measure of a procedure's ability to detect what is to be detected. It is calculated as the total number of seized or failed units, divided by the total number of units arriving that should be seized or are failures.

Efficiency : Economic performance calculated as the *output* (*e.g.* number processed or amount seized) divided by the *input* (typically the cost in terms of effort, time, and money).

Entry : movement of a pest, pathogen or disease from a source area to a location in Australia, usually associated with an import pathway.

Establishment : A non-native species forming self-sustaining populations outside its natural habitat.

Failure : Vessels or goods that do not meet the AQIS inspection, regulation or documentation requirements.

Impact : Environmental, economic or social change, positive or negative, that can be measured qualitatively or quantitatively.

Intervention : Application of a specified set of quarantine activities to determine the status of goods of quarantine interest. The rate of intervention is measured as the proportion of relevant goods or vessels subjected to intervention.

Leakage : 1. The undetected movement of goods or vessels of quarantine concern through an intervention process, or 2. The (estimated) amount or rate of undetected movements.

Risk : The likelihood of a specific type of event, possibly within a given time frame.

Spread : An increase in abundance and/or distribution following successful establishment.

2.3 Background

The objectives of this project are to review current data management and analysis processes that support import clearance; identify key issues in acquiring and analysing data; and develop guidelines that offer potential to improve the investment of resource in various import clearance surveillance, monitoring and auditing activities.

The role of AQIS is to reduce the risk to Australia's agricultural industries and the environment of an incursion by exotic pests and diseases. The primary mechanism by which this is achieved is quarantine control at the national borders, covering the various pathways by which goods may enter the country. This quarantine control is supported by an infrastructure of detailed information gathering and the physical inspection of imported goods. Considerable opportunities may exist within the infrastructure for enhancing or modifying current risk management practices. Data collection, analysis and reporting strategies may provide advantages in anticipating change and managing resources to minimise the risk of new incursions. This project will evaluate the potential for enhancing or modifying risk-based approaches to inspection and surveillance for some types of cargo.

"AQIS is Australia's first line of defence, reducing the risk to our agriculture industries and environment against exotic pests and diseases. Quarantine controls at Australia's borders minimise the risk of exotic pests and diseases entering Australia and protects our $32 billion agriculture export industries as well as our environment, tourism industries and lifestyle." (`http://www.daff.gov.au/aqis`)

In 2001 the Federal Government strengthened Australia's border security by funding a range of increased Quarantine measures in response to the outbreak of FMD in the UK and Europe. Within AQIS Import Clearance operations, this included measures for 100% intervention on the exterior surfaces of Sea Freight containers and Unit Loading Devices (ULD's – Air Cans), as well as with High Volume Low Value (HVLV) reportable documents. The Government also mandated that AQIS report on achieving these along with effectiveness targets of 96% for each function.

AQIS quarantine operations cover several pathways through which goods may enter Australia. The two pathways of entry that we reviewed are:

1. Air cargo, which comprises several classes of goods:
   (a) Unaccompanied person effects (UPE, covered by B534), which require manual assessment directly from the Cargo Report (CR) received from the Customs Integrated Cargo System (ICS);
   (b) Self-assessed clearance (SAC) goods, which are valued at $<$ \$1000;
   (c) Full import declaration (FID) goods, which are valued at $>$ \$1000; and
   (d) Reportable Documents (also called HVLV).
2. Sea cargo, which comprises the same classes of goods as above, but in practical terms is mainly cargo subjected to FID. Sea cargo also has regulations for the containers: consignments that are a *Full Container Load* (FCL and FCX) may be delivered as an intact unit to the importer (if the delivery premises are approved); a *less than a full container load* (LCL) must be unpacked at a suitable Quarantine Approved Premise (QAP) before delivery to the importers.

The external surfaces of air containers and sea containers are also inspected after unloading from the vessel, the sea containers under the External Container Inspection Regime (ECIR).

AQIS procedures also exist for areas such as air passengers, international mail, empty containers, and bulk cargo (*e.g.* bulk grain, vehicles, agricultural machinery). These are out of scope for this project.

2.4 Objectives and deliverables

The broad objectives of the project, as outlined in the project proposal, were to

- evaluate the potential for enhancing or modifying risk-based approaches to surveillance for some types of cargo and packaging
- outline the requirements for developing reporting systems to support these approaches
- describe systems for data acquisition, maintenance, analysis and reporting in quarantine systems that may assist AQIS to better target resources to respond to emerging risks and changing patterns of trade
- advise AQIS Cargo Management Group on the development of a framework to assess risk, improve data collection and analysis aligned with business processes, consolidate existing data and allocate resources to improve risk management

This project reviewed systems and data management protocols. It involved detailed discussions between ACERA researchers and AQIS Cargo Management staff. The discussions took place in Melbourne, Canberra and Brisbane. Based on this advice, we outline the needs of operational personnel and how AQIS goals may be achieved with current and future data acquisition, management and analysis systems.

Deliverable 1 : A review of current risk management strategies, supporting data collection requirements and processes (chapter 3). This outcome was achieved following;

- A familiarisation tour of AQIS operations for air and sea cargo
- An assessment of the information that AQIS currently collects to manage both commodity and non-commodity quarantine risk.
- An assessment of analysis processes in place for current information.

Deliverable 2 : Develop a risk framework that will assist the Import Clearance Program to identify significant quarantine risks and to develop processes and prioritise activities to treat those risks. The framework recommends, among other things, data collection processes and analysis processes to support prioritising the Program's activities. The framework and recommendations for analysis processes are presented in Chapter 4; recommendations for data collection processes are presented in Chapter 6, following the practical case studies in Chapter 5. The framework report;

- recommends processes for analysis of existing and potential information to manage quarantine risk (Chapter 4).
- recommends the information AQIS could collect to improve the management of quarantine risk for both commodity and non-commodity risks – *i.e.* the information needs to provide AQIS with statistical confidence that the quarantine risk of a commodity or non-commodity is acceptably low (Chapter 6).

Deliverable 3 : Developed and tested tools that are easily implemented with existing software and technology, that may assist officers in Canberra and regional settings to set priorities, view data, interpret trends, and make decisions about the allocation of effort and resources to better target areas of biosecurity risk. The tools are presented in Chapter 4. These tools were demonstrated and evaluated using selected case studies (Chapter 5) to illustrate the potential scope of risk-based approaches. The case studies demonstrate some of the advances that might be achieved under various scenarios.

2.5 Methods

The Victorian and Queensland regional field offices and the national office were visited. During three days at each location, port facilities and QAPs were inspected, and more than thirty AQIS officers and interested parties were interviewed. During the course of this project, AQIS officers outlined their work processes and data handling procedures, provided reports and examples of decisions. They discussed the limitations of current systems and their aspirations for improvements in information flows and decision support tools.

Once this information was compiled, the research group (the authors of this report) met and discussed potential case studies and the report structure. A subset of case studies was agreed, relevant to current operations and with the potential to illustrate operational effectiveness. An outline of the report was then drafted including a risk framework to support the implementation of surveillance and auditing innovations.

The draft report was circulated among the research group for comment and discussion. The team explored alternative formulations of the decision problems and evaluated their implications for inspection systems. Candidate solutions were applied to five case studies including

- Air Can surveillance (Section 5.1)
- Polished white rice (Section 5.2)
- Australian Fumigation Accreditation Scheme (AFAS) (Section 5.3)
- Air Cargo (HVLV/CAA) (Section 5.4)
- Import Clearance Evaluation (ICE) (Section 5.5)

These analyses were circulated among the research group for comment and revision. The draft report was then circulated more widely for critical peer review. Comments were incorporated into the final draft of the report.

3

Review of Current Practice

The two entry points for paperwork to obtain quarantine release for Sea and Air cargo are the front counter (FC) and the air cargo unit (ACU), respectively.

3.1 Front Counter

FC comprises both the AQIS staff at the physical counter itself and the operators who clear electronic applications from brokers, using Zetafax and email software.

A Full Import Declaration (FID) is lodged with the Customs system (ICS). The FIDs are then filtered by ICS using AQIS-supplied filters and hits only are forwarded to AQIS. The filters (or profiles) are based on a classification tree, and focus on commodity (*i.e.* the goods) and non-commodity (*i.e.* the pallets, packaging, *etc.*) characteristics. Commodity profiling is based on a commodity description tariff code as well as the identities of the brokers, suppliers, and importers. The commodity has a risk profile based on advice from Biosecurity Australia[1] (BA). The risk profile is obtained from the import conditions database, ICON, which sets out the conditions for certain goods to be imported, *e.g.* identifies the essential documentation. ICON includes both operational risk assessment advice and AQIS guidelines. The specification of attributes in ICON is determined in Canberra.

Non-commodity profiling is based on having correct treatment certification, packing declaration, and delivery (unpacking) location (metro/fringe-rural/rural). If non-commodity and commodity profiles show *clear*, then the goods are released. Non-cleared FIDs may be released or inspected, based on their documentation. Tiers of inspection are outlined below. Released goods are randomly subjected to audit, under the Broker Accreditation Scheme (BAS, see Section 3.4.3 below).

3.2 Air Cargo Unit

The ACU deals with reportable documents (HVLV), and air freight, from ICS. The ACU conducts inspections and document assessments at courier facilities including those of the companies DHL, TNT, FedEx and UPS. The ACU deals primarily with the reportable docs and Self-Assessed Clearance (SAC) entry types. Processing comprises a two-tiered screening process:

1. Firstly, the goods description (or the consignee or consignor information) is profiled from ICS. All hits against profiles are reported. These hits are processed on screen by an AQIS officer and the goods are either cleared or sent for further screening.
2. For SAC, the flagged records have paperwork inspected by an AQIS officer, and are either cleared, or inspected, or passed to the FC via AIMS for further assessment, *e.g.* if biological in nature.

[1] `http://www.daff.gov.au/ba`

If the AQIS officer deems it possible that the goods can be cleared by inspection, then the goods will be inspected. Of HVLV, all goods identified as reportable documents, and between 30% and 90% of goods identified as non-documents, depending on the country of origin and the flight, are x-rayed by ACS, possibly with an AQIS officer present.

The ICS profile database is centralized (located in Canberra, and maintained in NSW). New ICS profiles can be created by AQIS officers to cover new identified risks. Such profiles are applied immediately, and nationally.

3.3 Freight Inspection / Sea Cargo

Freight inspection covers commodity and non-commodity concerns as noted above. Depending on the product, officers will make a visual inspection or a take representative sample. Failures are recorded in AIMS and are used to construct AQIS profiles.

Inspection has three possible levels of detail:

1. tailgate inspection,
2. inspection of goods that have been fully or partially unpacked,
3. fully supervised unpacking of sealed container, and full inspection.

Full inspection comprises an examination of the non-commodity infrastructure and a representative sample or census of the commodity. This involves sampling according to standards reported in ICON. These standards originate with BA and are informed by and reflect the International Plant Protection Convention (IPPC) and the International Seed Testing Association (ISTA) guidelines, in the case of fruit and seeds respectively.

For example, paw-paw requires a sample of 600 fruit, with 450 examined under $\times 10$ magnification and 60 under $\times 100$ magnification. The boxes from which the fruit are selected are chosen representatively with some guidelines.

The number of boxes from which the goods are selected is open to inspector discretion and may informally reflect experience with a commodity or an importing region. Any insects or signs of disease found are sent to the AQIS entomologists and pathologists. Data are stored in the interception (Incidents) database. Samples typically proceed to one of three endpoints, which are not mutually exclusive:

1. Clear
2. Post-quarantine inspection
3. Detection: if there is a detection, the sample goes to the Operational Science Program (OSP) where it may be confirmed as a pest or disease, or it may be benign and the consignment is clear.

Inspection protocol (ICON) appendices specify the details of sampling. They are determined by ISTA and IPPC, but also by input from National Working Groups.

Non-compliance results in the establishment of a supplier profile, at which point the next 5 imports must be delivered to a class 1 QAP for inspection regardless of their content. The next 5 must be clear or the profile will remain.

In addition, containers to non-metropolitan destinations have a tailgate inspection at a QAP for insects, product description, seeds, dirt, packaging integrity regardless of profiling. If they fail then further treatment may be necessary, or a full inspection.

Data and systems assessments have begun. An internal audit review of AQIS profiling has been underway, and an internal data review project has commenced within the Import Clearance Program in Canberra.

3.3.1 Off-shore Inspection

Occasionally AQIS officers will be sent overseas to perform offshore inspection operations, that is, quarantine clearing before import. Offshore inspection is advantageous to AQIS because the risk is maintained offshore, and advantageous to the importer because a much more flexible response is possible if problems occur; *e.g.* local treatment followed by resubmission, or switching to local markets. Inspections of offshore-inspected goods are still necessary after arrival, to verify the off-shore process.

3.4 Industry Partnerships Unit

3.4.1 Quarantine Approved Premises (QAP)

All facilities at which imported goods are regularly received, stored, or examined, come under AQIS purview. Such facilities are called Quarantine Approved Premises (QAPs). The 2700 or so QAPs are divided into nine classes. Information about QAPs is stored in the Quarantine Premises Register (QPR), which controls the options that are made available to AIMS. The QAPs are scheduled to be inspected twice per year. The failure rate is low.

3.4.2 Compliance Arrangements

The Industry Partnership Unit also allows low-risk operations to be performed by external companies, in a set of schemes, using compliance agreements (CAs). Examples are: pig meat scheme, and onshore fumigation scheme. About 2500 such agreements exist across Australia. The Industry Partnership Unit maintains a physical record of these schemes.

3.4.3 Broker Accreditation Scheme (BAS)

BAS is a scheme whereby trained and accredited brokers or importers of record can assess and act upon commodity and/or non-commodity concerns without direct AQIS oversight in specific circumstances. BAS is controlled by the IPU unit in Canberra. The broker-assessed Automatic Entry Processing (AEP, for commodities) is a part of BAS, employed for low risk commodities such as NZ dairy and fin fish, and used vehicles. Information about auditing of the BAS scheme is presented in Section 3.9.

3.5 Off-shore Development Unit

AQIS has adopted the policy of keeping quarantine risks offshore as much as possible. The Off-shore Development Unit (ODU) is responsible for further development of the quarantine continuum outside the national borders by implementing and managing pre-border schemes. This includes the Australian Fumigation Accreditation Scheme (AFAS), Canadian Accredited Timber Scheme, Ethylene Oxide offshore treatment provider scheme and the Gamma Irradiation Offshore Treatment Providers Scheme.

The management of some of these schemes requires auditing of responsible overseas agencies and their service providers. Auditing for the AFAS scheme is further examined as a case study in Section 5.3.

3.6 Electronic Systems

Electronic Systems maintain the AQIS databases, including AIMS.

3.7 Business Planning and Intelligence

The Business Planning and Intelligence unit develops processes for setting budgets, charges to clients, and so on. The risks associated with these operations include the consequences of budget and resource shortfalls.

3.8 Business Improvement

AQIS IC is in the process of upgrading data collection and handling systems. The Business Improvement section is developing a new version of ICON, and developing and implementing the Strategic Information Management Plan (SIMP), which will focus on an improved AQIS import clearance system. Projected improvements include better data integrity, linkages between data layers, interface, and grain of incident reporting.

3.9 AQIS Auditing Infrastructure

BAS operations are audited with a minimum 3% sample by documentation inspection, selected through AIMS. If infractions are detected then the sampling rate for the broker is increased.

The ICE program provides a framework for evaluating import clearance effectiveness and an audit of quarantine operations. ICE includes the inspection of a randomly selected sub-sample of the 3% sample taken to monitor BAS. If an infraction is discovered, ICE issues a profile against the supplier; this feeds into the AQIS profile. Brokers may be penalized, depending on the nature of the breach. ICE data are stored centrally in Canberra.

As well, each Co-reg scheme has its own audit frequency, based on the associated risk, which is decided by Canberra, *e.g.* the pig meat scheme is audited 3 times per year. The audit uses a compliance checklist, including infrastructure, facilities for containment, etc.

Leakage surveys (the re-inspection of a sample of goods before their release to the client) are done to estimate the effectiveness of intervention for HVLV items, the external inspection of ULD, and other AQIS programs outside Import Clearance. Also, the verification of ECIR is available using tailgate inspection done for containers whose unpacking location is non-metropolitan.

4

Risk Framework

The broad objective of Import Clearance procedures is to maximise quarantine integrity. This involves allocating Australia's quarantine detection and inspection resources efficiently (ANAO, 2001, p. 20). Some of the criteria that contribute to the integrity of the system include:

- maximising efficiency and effectiveness
  - within commodities
  - within programs
  - over the Import Clearance system as a whole
  - focusing on high-risk items

In addition to interceptions, the roles of surveillance and inspections are to

- act as a deterrent to those planning to circumvent quarantine prescriptions,
- provide feedback on leakage rates associated with different pathways.

Data on incidents and leakage rates may identify areas where risks are relatively high. In general, efficiency and effectiveness will be enhanced by reducing inspections in areas where the risks are smallest and increasing inspections where the risks are highest. These demands are counter-balanced by the need for deterrence and to monitor trends associated with various pathways.

The risk management framework outlined below addresses in part the need identified by The Auditor-General's reports on managing quarantine effectiveness (ANAO, 2001, 2005). ANAO (2001) noted that there was only a limited risk-based process for allocating resources across different modes of entry and the quarantine continuum (p. 15), and that the value of border effectiveness measures did not address the potential consequences of entry (p. 16, 57).

ANAO (2001) recommended the development and deployment of tools to assist in the allocation of resources that are sensitive to the likelihood and consequences of incursions (ANAO, 2001, p. 20, 26, 35, 36). The need for systematic assessment of likelihoods and consequences across programs to support resource allocation and operational risk management was reiterated in the follow-up report (ANAO, 2005, p. 21, 58, 59, 66). ANAO (2001) recommended that resource efficiency should be considered across programs and the biosecurity continuum (figure 4.1).

4.1 Profiling Overview

In the context of quarantine, a profile is a set of characteristics that is used to identify where a possible threat is more likely to arrive and to establish the type of intervention required.

For some programs, comparing items arriving at the border with the list of profiles — profiling — is the first step in deciding what type of intervention (if any) is required. For other programs,

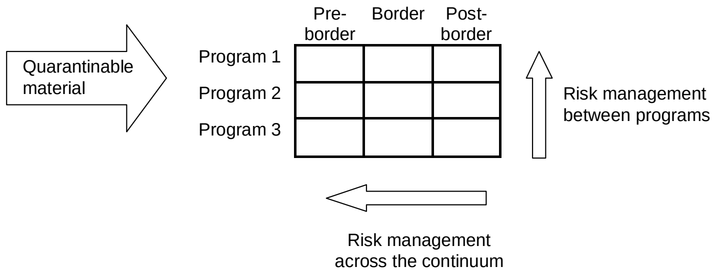

Figure 4.1: Risk management matrix: border programs and the quarantine continuum (Source: ANAO, 2001).

because of 100% intervention, profiling is only used for some components of the intervention or to provide staff with guidance about likely changes to types of approaching material.

Profiles are based on specific quarantine risks identified through import risk analyses and information provided by overseas sources (including advice from external agencies such as Biosecurity Australia and the World Health Organisation), experience gained by analysing data collected during the process, and data from leakage surveys. This analysis can also be used to determine if intervention techniques need to be changed to improve effectiveness or efficiency. Evaluation of items for which there were no interventions can be used to refine profiles.

Figure 4.1 outlines the relationship between profiling, intervention and evaluation. Items that are identified by a profile may be inspected or treated. These actions are evaluated, feeding back to improve the profiling system.

AQIS implements the profiling methodology across various programs. It delivers consistency in profiling processes and practices where possible through:

- establishing national and cross-program consistency in the methodology for developing and applying profiles;
- conducting regular reviews of the application and appropriateness of profiles, and encouraging cross-program exchange of methods and practices;
- training relevant staff in profile creation and implementation, and methods for describing profiles to internal and external stakeholders; and
- integrating profiling into business planning and maintaining awareness of changing operational conditions by ensuring that profiles evolve over time.

4.2 A tiered risk management system for the AQIS Operational Framework

This section of the report outlines a tiered risk management system that may be developed to provide a feedback between field observations and inspection protocols. For each commodity or activity, AQIS could apply the protocol summarised in figure 4.2, and the frameworks outlined below depend on its effective implementation. When resources are allocated at a National level, managers could decide on the relative risks posed by different commodities and activities, and allocate resources accordingly. The aim of decisions about the allocation of resources between commodities and activities would be to minimise strategic (business-wide) risks.

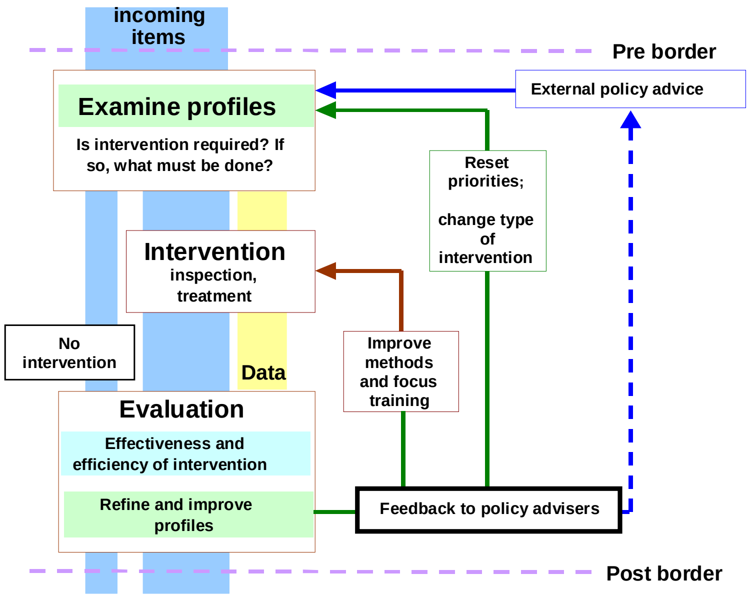

Figure 4.2: The ideal relationship between profiling, intervention and evaluation. This study creates tools to support the 'Feedback to policy advisers'.

The risk management framework should embrace existing business systems and support resource allocation to minimise strategic risks. Risk management at the National level requires structures that support strategic risk management. Risk management at the Regional scale requires flexibility to respond to seasonal variations in risk and the constraints that arise through uneven work flow and short-term contingencies. Thus, the risk management matrix in figure 4.1 represents a view that will be applied differently, at National and Regional scales.

This report outlines tiers of analysis that employ a consistent framework to accommodate these elements of risk management. Specifically, the risk framework provides a platform for managers at both National and Regional levels to make risk based decisions through analysis of the evidence on risk and performance effectiveness. The broad goal at both management levels is to maximise the benefits of investments in quarantine and inspection.

4.3 Tier 1. National-level strategic risk management.

At the National level, a primary objective is to set priorities among competing issues (commodities and activities), so that the highest risk issues are clearly identified and attract the closest attention. The system also needs to retain the benefits of learning about changes in risks over time, and providing a measure of deterrence against deliberate attempts to circumvent the system.

Figure 4.3 describes a template for evaluating relative likelihoods of entry, consequences of entry, and the cost and feasibility of inspection systems and management intervention options. Consequence assessments consider the extent of impacts (through spatial analysis) and the kind and intensity of impacts within the extent of occurrence (evaluated through consideration of economic, social and environmental criteria). A values hierarchy may be used to structure the representation of the elements at risk from the invasion of the pests, diseases or pathogens under consideration. The assessments of likelihood of entry, establishment and spread, and consequence assessments would typically be provided by Biosecurity Australia.

Multicriteria analysis may provide an appropriate platform for the "Decision analysis" step in figure 4.3. AQIS evaluates the relative risks of the suite of commodities and activities it manages. Commodities or activities that pose very low likelihood of entry or have very low consequences are placed low on the list of national priorities. Commodities or activities that rank relatively high under likelihood and consequence are ranked highly on the list of national priorities. Ranks take into consideration economic, environmental and social risk. Given that not all consignments and not all lots of commodities can be examined, AQIS is obliged to make decisions about the allocation of resources for its inspection activities. Some activities are prescribed as mandatory under the IRA, and an ability to provide feedback, for example to Biosecurity Australia, would be valuable in these cases.

Deliberations that use multicriteria decision analysis allow managers to make decisions based on structured assessments of the important contributing factors. The factors may be combined using various rules, and criteria may be weighted by their relative importance. The sensitivity of decisions can be evaluated by omitting criteria or by changing their relative importance.

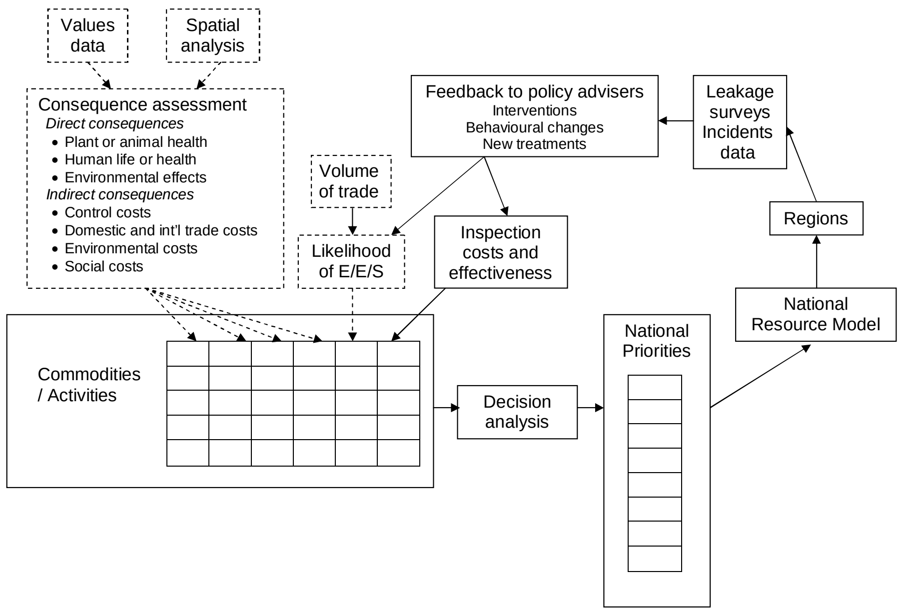

Figure 4.3: National-level risk management framework. Activities compete for resources. Biosecurity Australia provides information on the likelihoods and consequences of entry, establishment and spread and recommendations for treatment (indicated by the dashed lines). AQIS evaluates the relative risks of the range of commodities and activities within its jurisdiction and allocates resources to activities accordingly.

The framework outlined in figure 4.3 translates evaluation data into a system to support efficient allocation of resources, where the flexibility to do so exists. It creates an opportunity for feedback to encourage continuous improvement. The framework in figure 4.3 is established in a spreadsheet attached to this report, including tables showing values considered in biosecurity risk assessments, a weighting scheme for these values and an aggregation of the values into a single measure of consequence. This report recommends that an implementation program include tests of alternative value trees, weights and aggregation methods, with the ultimate aim of finalizing a tool that could be used in risk minimization and feedback systems.

The system outlined in figure 4.3 may provide a basis for more transparent and repeatable decisions. It makes use of likelihood and consequence information provided by Biosecurity Australia, and takes into account the relative effectiveness and costs of inspections.

4.4 Tier 2. Regional-level tactical risk management

Regional-level deliberations involve allocating resources among competing demands from inspection and surveillance activities across a wide range of activities. National priorities play a part in determining these priorities. In addition, each region confronts unique sets of commodities, trade conditions, volumes of material, staff, and other operational constraints and demands (figure 4.4).

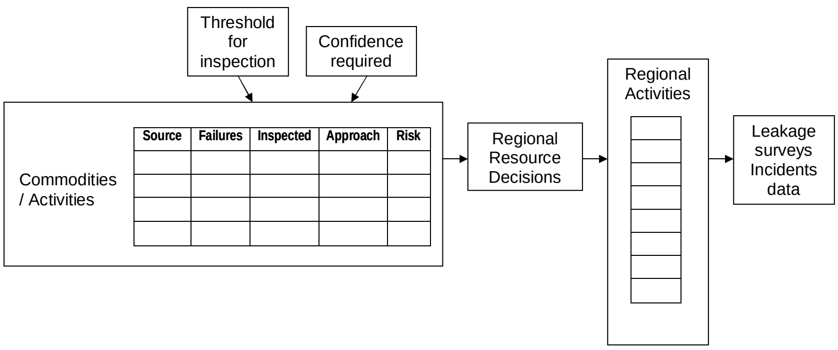

Figure 4.4: Regional-level risk management framework, showing the input data and conditions (thresholds and confidence levels) required to make decisions about allocating inspection resources among commodities and activities.

Inspection systems are affected by regional operational conditions and the specifics of commodities and trading conditions experienced at local ports (including the time of year, volume of traffic and trade). Criteria that contribute to local priorities include short-term trade-offs among commodities and activities and the availability of appropriately trained staff. They may be combined using the structured decision analysis, with different criteria and weights that reflect regional contingencies.

Risk analysis is outlined below. The proposal describes a method for assessing the relative risk of inspection regimes that account for the volume of trade in a specified period (a day, a week, or a year), the likelihood that each unit of trade contains a threat (the relative riskiness of commodities, sources or activities), the costs of inspection and consequences of failing to detect a failure.

4.5 Risk analysis

We outline a strategy for the risk-sensitive allocation of resources to inspection and auditing. The case studies that follow will build on this platform in different directions.

We emphasize that the risk analysis strategy outlined below is modular. Each of the overarching themes of defining risk, predicting risk, and balancing risk, whilst interrelated, can be replaced by other candidate strategies.

The platform of our proposed approach to risk analysis is as follows. We suggest that accurate ongoing estimation of the failure rate for a commodity is an important step in identifying areas of high and low risk. We blend the joint challenges of detection and estimation into a single index, called *risk*, which is defined below. We propose the risk index as a means of profiling commodities, that is, identifying *high-risk* commodities for 100% inspection, and for allocating resources to the monitoring of *low-risk* commodities. Generally speaking, this approach will allocate more resources to the monitoring of commodities that are (a) known to be high-risk, or (b) whose risk is poorly known.

AQIS has three main resource allocation challenges:

1. tactical, short-term resource allocation, for example: an officer has spare time for inspection: what should they inspect? Or, an officer has to make a choice between inspecting two consignments, and there is insufficient time to inspect both; should they choose one and if so which should they inspect? Or should they inspect both superficially?
2. strategic, long-term resource allocation, for example: what is the expected cost and benefit of inspecting commodity x at 100%? Do the historical data support such a policy? Could a flexible profiling approach be adopted, which might maintain high detection rates whilst reducing intervention rates?
3. frequency of auditing, often in the context of limited resource availability. This challenge combines characteristics of the previous two; the resource limitation of the first combined with the strategic monitoring considerations of the second.

The first challenge is subject to resource shortage, in that we assume that inspection of consignment A precludes the possibility of inspecting consignment B. This tactical challenge is resolved internally, within AQIS.

The second challenge is not, strictly speaking, subject to resource shortage because AQIS inspections are run on a cost-recovery basis. However, this strategic challenge involves relationships with other stakeholders, such as Biosecurity Australia, and AQIS may well have no direct, short-term leeway in its resolution.

We briefly sketch a solution below that provides a risk-sensitive approach to resolving these challenges.

1. Every commodity is assigned a *predicted risk*, $\hat{f}$. This is an upper bound on the probability (defined below) that an individual unit of the commodity contains at least one failure.
   (a) This risk can be defined in many ways; we choose the upper limit of the one-tailed 95% confidence interval.
      i. This interval can be computed in many ways; we presently recommend Jeffrey intervals.
      ii. Model-based empirical Bayes intervals might have better statistical properties, but this consideration is outside the scope of the current exercise.
   (b) The risk can be updated every so often (month, season, year) to reflect new information, whether that be fresh inspection data, or evidence of heightened risk from other sources.
2. Short-term, tactical resource allocation is based purely on the predicted risk: given a choice between inspecting a number of consignments, an officer should inspect the riskiest consignment.
3. Long-term, strategic resource allocation proceeds as follows.
   (a) AQIS nominates a risk cutoff r, *e.g.* 1%, possibly with input from other stakeholders, and possibly varying across commodities.
   (b) Every commodity with $\hat{f} > r$ is inspected at 100%
   (c) Every commodity with $\hat{f} \leq r$ is inspected at a sufficient rate such that the probability that the observed risk will turn out to be, in fact, higher than r is, *e.g.*, 5%[1]. That is, we

[1]Alternatively, higher than $r - \delta$, where δ is a buffer

inspect sufficiently frequently that we can assure ourselves with defined confidence that the true failure risk is less than the nominated cut-off, unless the underlying failure rate changes in the intervening period. We then use the resultant monitoring data to reassess the prediction at the next time step.

i. The rationale behind the monitoring rate is to be reasonably certain that, if the failure rate for the commodity does not increase, then the commodity will be assigned to monitoring again in the future, whereas if the failure rate increases sufficiently, then the algorithm will recommend 100% inspection.
ii. The inspection rate is calculated by divided the necessary sample size by the estimated number of consignments arriving. See below for computation details.
iii. In cases of allocation under constraint, whether that be a limit to the number of consignments that can be inspected, or a limit to the rate of resource allocation, r can be varied to match the available resources.

4.5.1 Details

Consider two commodities, A and B. Different volumes of trade are expected over some fixed period (a week, a month or a year).

We will denote the *predicted* volume of trade (the number of units) expected of commodity A as N_A, and the predicted *risk* that an individual unit of A will contain at least one failure will be denoted f_A. The risk will be related to the predicted probability of failure of a consignment of commodity A, inflated to reflect uncertainty about the quality of the prediction. We define and explain f_A below.

Predicting risk rates

In order to allocate the resources to inspection of the different commodities it is necessary to predict the volume of trade and the risk rate for each commodity. Estimation and monitoring of the approach rate of failures is essential for a risk-sensitive allocation process.

These predictions will generally come from data collected in the previous year. That is, we are forced to use the failure rate and the approach rate from the previous year as predictors for the failure rate and the approach rate of the year for which the allocation will be made. Hence our nomination of the two rates for the purposes of allocation will be subject to uncertainty, even if the estimates for the previous year are exact for the previous year, as they may be[2]. It is also possible that data collected in other earlier years could also be deployed to get a better prediction.

Therefore, the predictions of N_A and f_A will generally be subject to error. To simplify the problem of allocating resources we will assume that N_A is relatively accurately predicted (typically as a percentage increase in volume for the year), and that therefore the overwhelming contribution of uncertainty comes from the failure rate.

Assume that we have historical monitoring data for each commodity. We assume that the number of failures detected for each commodity is a binomially distributed random variable. That is, let k_A denote the observed number of failures of commodity A, then

$$k_A \sim \text{Binomial}(p_A, n_A) \tag{4.1}$$

where n_A is the number of consignments inspected of commodity A, and p_A is the unknown probability of failure. With this model in place, we can proceed to predict f_A, the risk.

[2]In statistical terms this is called a super-population model; we are trying to estimate the uncertainty related to an underlying process.

Paying for uncertainty

We argue that we need to consider both the estimate of the failure probability p_A and the accuracy of the estimate when allocating resources. Given two commodities with the same failure rate but different inspection and approach rates, we would be *less certain* about the estimate of the failure rate for the commodity that has the lower approach rate. A risk-sensitive solution requires that we take account of the uncertainty of the estimate in our allocation, effectively forcing us to pay for higher uncertainty in the same way that we pay for higher failure rate; that is, by increasing the allocation.

An integrated way to achieve a risk-sensitive estimate is to use a *quantile*, a one-tailed confidence interval, to predict the risk rate, instead of using the failure probability. That is, where we would ordinarily choose the value that is the best supported by the data, instead we will choose a value that represents an interval that has a specific designed probability of including the true rate under repeated sampling. For example, we might choose a failure risk that corresponds to the upper 95% confidence interval of the true failure rate of the underlying process. *Informally*, we could describe this value as a limit that we are 95% confident that the true failure rate is *below*. See figure 4.5 for a demonstration.

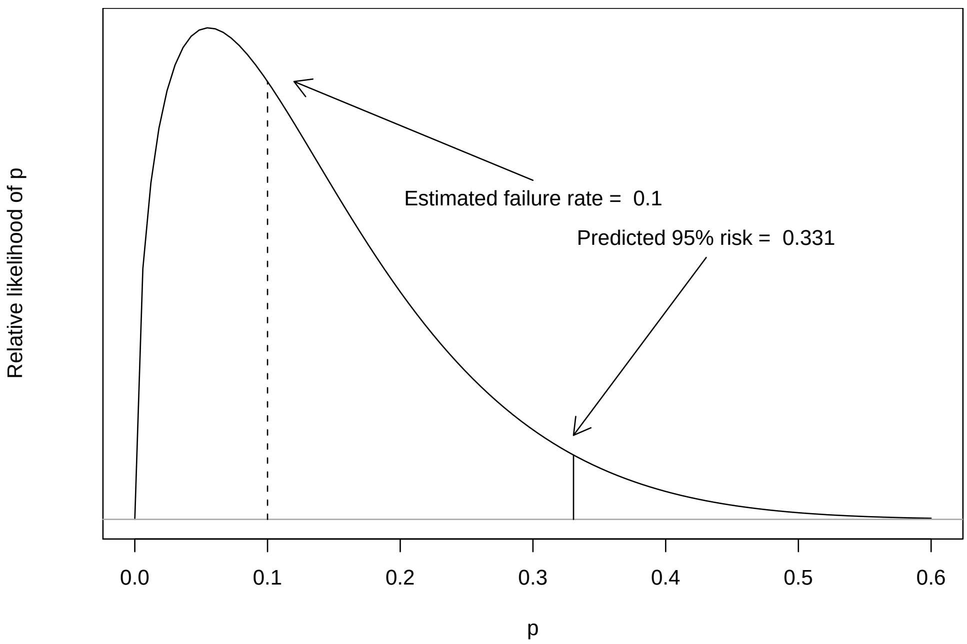

Figure 4.5: An assumed underlying probability density function of the failure rate for commodity X, showing the estimate of the failure rate and an estimate of the 95% failure risk. The probability density function is the Beta distribution.

The actual quantile nominated will generally depend on how much data are available, and how the analyst wishes to value uncertainty about the true rate.

Estimating the uncertainty-inflated failure risks

We anticipate that the estimated failure rate will usually be close to zero, and that there will be very few or zero failures for some commodities. We wish to obtain an upper one-tailed confidence interval for this rate; we will use the upper limit of the confidence interval as a risk rate.

The behaviour of different confidence intervals for the binomial rate has been the subject of some research (see, *e.g.*, Madden et al., 1996; Brown et al., 2001; Cai, 2005). For one-tailed confidence intervals with good properties close to 0 and 1, Cai (2005) recommends *Jeffrey intervals*. Jeffrey intervals are Bayesian intervals, with a prior distribution on the binomial parameter equal to Beta$(0.5, 0.5)$. Jeffrey intervals do not guarantee $(1-\alpha)\%$ coverage but they represent a good compromise. If absolute coverage is necessary then a preferred alternative would be Clopper-Pearson intervals which do guarantee $(1-\alpha)\%$ coverage but are also known to be conservative.

Jeffrey intervals are easily computable in standard spreadsheets, such as Microsoft Excel[3]. Briefly, the upper one-tailed confidence interval is computed from the inverse of the cumulative density function of the Beta distribution. The Beta distribution requires two parameters; for our application they are computed as $a = k_A + 0.5$ and $b = n_A - k_A + 0.5$. Thus if we wished to use an upper 95% confidence interval then the function call in Microsoft Excel for f_A, the estimated 95% upper confidence interval for the unknown rate, would be:

$$= \text{BETAINV}(0.95, k_A + 0.5, n_A - k_A + 0.5) \tag{4.2}$$

where k_A is the observed number of failures for commodity A, and n_A is the number of consignments inspected of commodity A. Informally, equation 4.2 is

$$= \text{BETAINV}\,(0.95, \text{ number of failures} + 0.5, \text{ number of non-failures} + 0.5) \tag{4.3}$$

Then,

- if the predicted risk is greater than a pre-determined cutoff, then the risk-sensitive analysis recommends that the commodity should be inspected at 100%, whereas
- if the predicted risk is less than the cutoff, then the risk-sensitive analysis recommends that the commodity should only be sampled with a view to *monitoring* the risk associated with the commodity.

These two tasks are very different. For inspections, examining 100% of the commodity is desirable. For monitoring, the desired sample size can be computed as a fixed number of consignments. That is, the outcome of this analysis might be a statement along the lines of "If you sample 500 widgets over the next year *and* if the failure rate for widgets does not increase from its current level then we are 95% confident that the estimate of the upper confidence limit for the failure rate of widgets at the end of the year will be below 1%". Note that the number of inspections needed for monitoring does *not* depend on the number of widgets arriving. This means that, following the example above, the number of widgets to inspect will be 500 whether the approach rate is 501 widgets per year or 10000 widgets per year. It is possible that the required number of samples may be greater than the estimated number of items arriving, in which all would be sampled, and the desired risk would not be possible to obtain even if all the commodity is inspected.

There are at least two ways of thinking about the problem of allocation. The intuitively satisfying approach is to select the sample size x that provides us with, say, 95% confidence that the true risk for the commodity will be below our cutoff. However, this approach opens up the possibility that a borderline commodity will flip back and forth from monitoring to 100% inspection. Such flipping may have unintended consequences, such as confusion for field officers[4]. Instead, we tentatively suggest to select sample size x as the value that provides us with, say, 95% confidence that the risk for the commodity at the end of the year will be below our cutoff, if the number of failures does not increase. Our suggestion amounts to placing risk boundaries on future outcomes of the underlying process, rather than the process itself. In statistical terms this approach is like constraining the prediction interval rather than the confidence interval.

[3] Clopper-Pearson intervals can be easily computed using the `BETAINV()` in Microsoft Excel, and indeed this forms the basis of the effectiveness calculations already performed by AQIS IC. The CP version of equation 4.2 replaces $k_A + 0.5$ by $k_A + 1$ and $n_A - k_A + 0.5$ by $n_A - k_A$ (see, *e.g.*, Krishnamoorthy, 2006)

[4] Although not in case that the system is entirely automated.

To find the preferred monitoring rate x, we must predict the upper limit of the number of failures that we would expect to get if we inspected x consignments and if the number of failures does not increase. Assume that using our historical data we have k failures in n inspections. Then, the 95% upper confidence limit for the future failure rate is computed using equation 4.2, with the following changes. The expected number of failures is now $x * k/n$ instead of k, and the number of inspections is (by definition) x instead of n, so, the predicted future 95% upper confidence limit for the predicted number of failures from a future sample of size x is:

$$k_{95} = \text{BETAINV}(0.95, x/n * k + 0.5, x - x/n * k + 0.5) * x \tag{4.4}$$

We now want to find the sample size x that will result in a predicted risk lower than our cutoff, if the failure rate is no greater than it has been in the past. We obtain that sample size by finding the value of x that makes the following equation true:

$$r = \text{BETAINV}(0.95, k_{95} + 0.5, x - k_{95} + 0.5) \tag{4.5}$$

where r is the nominated risk cutoff (*e.g.* 1%). x is an unknown in both equations 4.4 and 4.5, so they are both required for the solution.

In the case of allocation under resource constraints, we fix either the total number of inspections that can be done ($\sum x$), or the nominated sampling rate ($\sum x / \sum n$), and find the value for r that satisfies equation 4.5.

We note that the method suggested for choosing the sample size places considerable emphasis in the greater uncertainty that arises from small sample sizes. There are alternative ways of calculating a sample size that may cause more frequent switching between whether the group should be sampled at 100%, or at a lesser rate, but will also be less conservative in terms of penalizing the uncertainty.

The practicality of sampling also needs to be considered in some cases. Even if the 100% sample list remains relatively constant, the sample rate for the non-100%-sampled categories will change each year. This regular change may lead to complications that outweigh the benefits, especially where a human must make the decision rather than an electronic system. Under these circumstances using a constant rate (say 5% or 10%), would be more practical, perhaps with the proviso that if the number of items approaching is low, increase the sampling rate.

Example

Assume, without loss of generality, that we have previously been sampling at 100%, and that the monitoring data are as represented in Table 4.1.

Table 4.1: Example risk-sensitive allocation. k is the number of failures from n previous inspections out of N consignments. $\hat{f}$ is the predicted risk, and TFSR (%) is the tentative future sampling rate expressed as a percentage, targeting a 1% risk cutoff (see equation 4.5).

Commodity	k	n	N	$\hat{f}$	TFSR (%)
A	2	100	100	0.05425	100.0
B	20	10000	10000	0.00285	10.9

Short-Term Allocation: Allocate inspection efforts to A as it has the higher risk.

Long-Term Allocation:

1. The risk for commodity A is higher than 1%, therefore we propose to inspect A at 100%.

2. The risk for commodity B is less than 1%, therefore we propose to sample 1088 units of commodity B, which represents a predicted sampling rate of 10.9%. This is the sampling rate that will result in a less than 1% predicted risk with 95% confidence if the failure rate does not increase.

Generalizations of this example will help clarify the implications of our proposed strategy.

- For example, if the cutoff for risk were 0.5% instead of 1%, then:

    1. Commodity A is still inspected at 100%
    2. Commodity B is still sampled, but now we sample 4524 consignments, which leads to a predicted sampling rate of 45%. This is the sampling rate that will result in a predicted risk of less than 0.5% with 95% confidence if the failure rate does not increase.

- Also, if the *future* number of units of commodity B increases from 10000 to 100000, with the cutoff for risk remaining at 1%, and the failure rate stays the same at 0.2%, then we still sample 1088 units of commodity B, which leads to a predicted sampling rate of 1.1%. This is the sampling rate that will result in a predicted risk of less than 1% with 95% confidence if the failure rate does not increase. In short, *the number of consignments to inspect does not depend on the future approach rate*.

- Finally, if the *observed* number of units of commodity B increases from 10000 to 100000, with the cutoff for risk at 1%, and the failure rate for the commodity stays the same at 0.2%, then we still sample 1088 units of commodity B, with a predicted sampling rate of 0.011. This is the sampling rate that will result in a 1% predicted risk with 95% confidence if the failure rate does not increase. In short, *the number of consignments to inspect does not depend on the historical approach rate, only the predicted failure rate and its uncertainty.*

Resource-Constrained Long-Term Allocation: Continuing from the previous example, imagine that we are required to allocate resources between A and B subject to a constraint that we can only inspect a total of, say, 100 units. To allocate these 100 units between A and B we should determine the risk cutoff that corresponds to the sum of the sample sizes being 100. That is, we choose the risk cutoff that corresponds to the prescription of sampling exactly 100 units. The resources are then allocated according to the sample sizes. In this case, to achieve this goal we must sample 61 units from resource A and 39 units from resource B. The realized risk cutoff that leads to this allocation is 14% across both resources.

Finally, imagine that we are required to allocate resources between A and B subject to a constraint that we can only inspect, say, 5% of all possible units. To allocate the 5% between A and B we determine the risk cutoff that corresponds to the sampling rate being 0.05. The resources are then allocated according to the resulting sample sizes. To achieve this goal we must sample 100 units from commodity A and 405 units from commodity B. The realized risk cutoffs that lead to this allocation are 10% for commodity A and 1.9% for commodity B. These are different because commodity A is being censused (see Table 4.1).

4.5.2 Sensible deployment

We emphasize that the risk results from analyses like these should be treated as ball-park guidelines, rather than hard and fast rules. The practicalities of the physical constraints of inspection and auditing will inevitably over-weigh the risk analysis. For example, it may be that the risk associated with a particular commodity is extremely small. Nonetheless, it is likely to be important to maintain a level of surveillance for the purposes of monitoring and deterrence. In this case it will be worthwhile keeping aside a portion of the available resources for that purpose.

4.5.3 Multiple sources of risk

The risk-sensitive allocation generalises in a straightforward way to multiple sources. Inspection effort is weighted by risk (probability of a threat times volume of trade). It provides a surveillance element even when probabilities or volumes are very low, thereby providing continuing deterrence and information on rates, in case probabilities change unexpectedly, while focusing most resources in areas where risks are highest. Appendix B outlines some other formulations with different objectives that result in different (and in our opinion, less useful) resource allocation strategies.

4.5.4 Consequences of scale

Risk considerations will be made at a number of hierarchical scales. For example, the risk of a commodity might be computed at the overall level and also broken down by region. Due to the nature of the risk index that we have nominated, the risk will invariably be higher at the regional level than at the national level. This difference is, intuitively, because we have more information at the national level than at the regional levels, and our risk index penalizes lack of knowledge.

To provide a numerical example for monitoring, imagine that for a certain commodity we have only two regions, A and B. Table 4.2 provides the details.

Table 4.2: Example of risk-sensitive allocation across hierarchical scales. k is the number of failures from n previous inspections out of N consignments. $\hat{f}(\%)$ is the predicted risk as a percentage, and TFSR (%) is the tentative future sampling rate expressed as a percentage targeting 1% risk cutoff (see equation 4.5), and $\hat{n}$ is the projected sample size.

Region	k	n	N	$\hat{f}$ (%)	TFSR (%)	$\hat{n}$
A	20	10000	10000	0.285	10.88	1088
B	20	10000	10000	0.285	10.88	1088
National	40	20000	20000	0.257	5.44	1088

Notice that the risk identified at the regional levels is higher than that of the national level, and that the predicted sample sizes are identical at both scales. In other words, to achieve the same risk cutoffs within the regions as is desired at the national level, we would need to monitor twice as many samples.

A further challenge is that as the analysis drills deeper into the hierarchy of a dataset, sparsity affects the quality of estimates. We suggest and implement a modelling strategy that supports deeper drilling in Section 5.6.2.

5

Case Studies

In this chapter, we demonstrate the use of statistical risk analysis on a selection of case studies, using data kindly supplied by AQIS. Our goal is to use the case studies to illuminate our suggested approach to risk analysis. Further refinements will be possible in each case, and will almost surely be beneficial.

It is important to recognise that from the point of view of AQIS IC, it makes perfect sense to *initiate* a monitoring program for a potential non-conformity at a very high level. It is also reasonable, after some time has passed, to review the monitoring program, having learned more about the probability of occurrence of the non-conformity. The deployment of statistical risk analysis tools helps to answer the questions: how to go about such a review, what data are necessary, how many data are needed, what statistics are informative, and when the evidence is clear.

The case studies cover a range of realistic decision scenarios, and each uses the approach outlined in section 4.5 to a different degree. Loosely speaking, these scenarios can be categorized as follows.

- Changing the level of surveillance for a single commodity.

  For example, the outsides of aircans (Section 5.1) and consignments of white rice (Section 5.2) are both inspected at 100%. Surveillance data have been collected for several years now. How could we answer relevant questions such as: does AQIS need to do so much?

- Allocation of resources between a small number of options.

  For example, AFAS regularly audits fumigation companies in other countries (Section 5.3). There are practical limitations to how many countries the auditors can visit in any given year, and how many facilities can be audited within each country. How can we use the information available from previous auditing exercises, and possible other information, to help guide these choices?

  Likewise, inspection of reportable documents is required to be 100% (Section 5.4). When the number of arriving reportable documents is low, officers informally check incoming goods on Cargo Assurance Air (CAA, also called the 'freeline'). Anecdotal evidence suggests that many more failures are found in CAA than in reportable documents. How could AQIS officers balance their effort in a reasonable way?

- Allocation of resources between a large number of options.

  For example, the present selection of consignments for ICE auditing is random and balanced (Section 5.5). However, it might be possible to justify a reduction (or an increase) in the level of intervention for a given commodity or region (or some combination). How would such a change in the level of intervention be chosen, and what benefit might be expected?

As implied above, the case studies also span a range of AQIS functions, including inspection (aircans, Section 5.1; white rice, Section 5.2; and air cargo, Section 5.4) as well as auditing (AFAS, Section 5.3, and ICE, Section 5.5).

Finally, the case studies include scenarios for which resource allocation proceeds under resource (or other) constraints, specifically the AFAS (Section 5.3) and Imported Food (Section 5.6) case studies.

All analyses were performed using the open-source, free statistical environment R (R Development Core Team, 2008).

5.1 Aircan surveillance

5.1.1 Summary

The failure rate for aircan inspection for 2007 is 0.084%. The national failure risk for external aircan inspection for 2007 is 0.0918%. Note that the failure risk is an estimated upper confidence interval for the average rate.

The total number of failures detected nationally from July 2003 until December 2007 is 1711, from a total of 1,696,130 inspections, which is an average rate of 0.101%.

The regional and national predicted annual risks are presented in table 5.1 and figure 5.2. The highest regional failure risk for external aircan inspection for the year of 2007 is 1.5%, in the Far North region. The regional and national failure rates are presented in figure 5.1.

When interpreting the results in table 5.1, it is important to keep in mind that there are two contributing reasons to any region having relatively high failure risk: firstly, many failures may have been detected (*e.g.* thirty-three failures in the Far North in 2007, see table C.1), and, secondly, a relatively small number of aircans may have been available for inspection.

To simplify the explanation, we assume that all possible aircans have been inspected, and that the inspections are 100% effective.

Table 5.1: Estimated 95% failure risk, average failure rate, and tentative future sampling rate (TFSR) for external inspection of aircans, 2007, presented by region and nationally. The failed and inspected columns refer to the totals for each region in 2007. $\hat{f}_{0.95}$ (%) is the predicted risk, expressed as a percentage. $\hat{p}$ (%) is the average failure rate. TFSR (%) is the tentative future sampling rate, expressed as a percentage, for a risk cutoff of 1% (see equation 4.5). TAS and NT had no inspections in the period.

Region	Failed	Inspected	$\hat{f}_{0.95}$ (%)	$\hat{p}$ (%)	TFSR (%)
Brisbane	58	37743	0.190	0.154	2.46
Far North	33	2957	1.470	1.116	100.00
NSW	137	207764	0.076	0.066	0.33
SA	59	17510	0.415	0.337	10.10
VIC	24	91491	0.036	0.026	0.65
WA	0	14067	0.014	0.000	3.83
National	311	371532	0.092	0.084	0.20

Note that the recommended sampling rates at the national and regional levels are not necessarily commensurate. That is, aggregation of the regional recommended sampling rates will not result in the national sampling rate.

The results of this analysis have two potential applications.

1. The reported risk can be used to support an assessment of the utility of 100% inspections. Our results suggest that 100% inspection of aircans does not necessarily substantially reduce the risk relative to a lesser sampling rate, and that a risk-sensitive approach may recommend sampling aircans at a lower rate and focusing inspection resources elsewhere.

2. The results can be used to identify those regions that have higher failure risks. Table 5.1 identifies Brisbane (58 failures from 37743 inspections in 2007), South Australia (59/17510), and Far North (33/2957) as having higher failure risks than the other regions. (See tables C.1 and C.2 for failure and inspection counts by region and year.)

   On the strength of these results, it may be reasonable to maintain the current high inspection rates in these three higher risk regions, but to consider reducing the inspection rates for Victoria (24 failures from 91491 inspections in 2007), Western Australia (0/14067), and New South Wales (137/207764). In this way the reduction in inspection rates suggested in the previous step can be deployed with best effect.

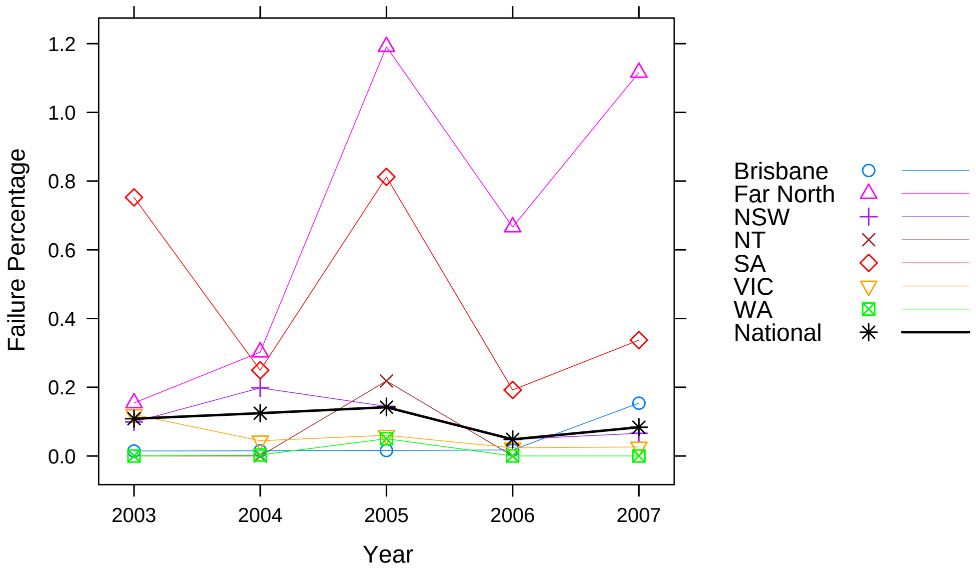

Figure 5.1: Annual failure rate for external inspection of aircans, by region and nationally, expressed as a percentage. The thick black line is the national line.

5.1.2 Analysis

AQIS policy specifies that all air cans will be inspected externally. In this context, a failure corresponds to there being contamination on the outside of the aircan. After numerous years of inspection, very few failures have been found. Here we estimate the failure risk, by region and nationally.

The aircan inspection data that we received from AQIS were monthly counts of inspections and failures, by region. To simplify the presentation we chose to construct summary statistics annually, both at the regional and national level. The national-level statistics are the unweighted aggregate of the regional statistics.

Our goal was to obtain a conservative indication of failure risk. Therefore we report the upper limit of the one-tailed 95% confidence interval for the estimated failure rate. We can interpret this figure as being a conservative upper limit for the true rate. Otherwise the analysis was performed as laid out in Section 4.5.

We converted the rates to percentages for our summary statistics, and for figure 5.2 for ease of interpretation.

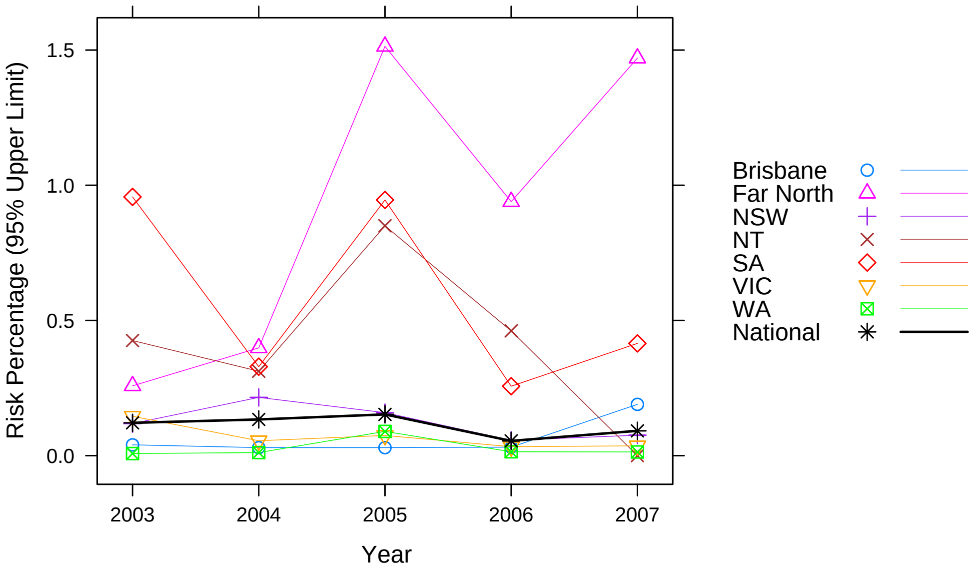

Figure 5.2: Estimated annual 95% failure risk for external inspection of aircans, by region and nationally, expressed as a percentage, for a risk cutoff of 1%. The thick black line is the national line.

5.1.3 Further directions

Our goal is not to provide the definitive risk analysis for aircan inspection, but rather to use the aircan case study to demonstrate the tools that we suggest will be useful for risk analysis. Given sufficient information there is no doubt that a better allocation could be constructed.

Dividing the shipments by destination is an artificial approach that we adopted because the data were readily available. It would make more sense to divide the shipments by source country, if that information were available.

The documented analysis treats each year separately, which prevents any sharing of information from year to year. Arguably, the division of collection periods into years is artificial, and a better technique would permit aggregation of data from more than one year, which would simplify the identification of seasonal differences within the year. This approach can be easily implemented by accumulating the observed failure and inspection counts across a number of years, possibly in a weighted fashion. The results of a version of this analysis are presented in Section C.1.

5.1.4 What have we learned?

1. AQIS has the necessary data to question whether 100% inspection of aircans is worthwhile.
2. The information from leakage surveys for aircraft inspections could be used to adjust the number of aircans that are of concern.

Information Needs

1. AQIS already aggregates inspection data for external inspection of aircans by month and year, and by the region to which the aircan is delivered. This dataset was used for the analyses reported above.

2. A more precise profiling approach would be possible if more information about the individual aircans were readily available. For example, failure rates might vary by source country, and profiles that were constructed with this knowledge would enable better-targeted surveillance.

5.2 White rice surveillance

5.2.1 Summary

The failure rate for white rice inspection for 2007 is 0%. The failure risk for white rice inspection for 2007 is 1.5%. Note that the failure risk is an estimated upper confidence interval for the average rate.

The total number of failures detected nationally from July 2003 until December 2007 is 3, from a total of 569 inspections, which is an average rate of 0.527%.

The regional and national annual failure risks are presented in table 5.2 and figure 5.4. The highest regional failure risk for white rice failure for the year of 2007 is 26%, in the Other region. When interpreting these results, it is important to keep in mind that there are two contributing reasons to a relatively high value: firstly, failures may have been detected, and, secondly, a relatively small number of white rice lots may have been available for inspection. Tables containing number of inspections and number of failures by region and year are presented in Section C.2. The regional and national failure rates are presented in figure 5.3.

To simplify the explanation, we assume for these estimates that all possible white rice consignments have been inspected, and that the inspections are 100% effective.

Table 5.2: Estimated 2007 95% failure risk and tentative future sampling rate (TFSR) for inspection of white rice, presented by region and nationally, expressed as a percentage. The failed and inspected columns refer to the totals for each region across all years. *Other* represents NT, QLD, SEQLD, and TAS combined. $\hat{f}_{0.95}$ (%) is the predicted risk, expressed as a percentage. $\hat{p}$ (%) is the average failure rate. TFSR is the tentative future sampling rate, expressed as a percentage, for a risk cutoff of 1% (see equation 4.5).

Region	Failed	Inspected	$\hat{f}_{0.95}$ (%)	$\hat{p}$ (%)	TFSR (%)
NSW	0	35	5.3	0	100
SA	0	12	14.5	0	100
VIC	0	39	4.8	0	100
WA	0	39	4.8	0	100
Other	0	6	26.4	0	100
National	0	131	1.5	0	100

Although the recommended sampling rates at the national and regional levels are commensurate in this example, they need not necessarily be. That is, aggregation of the regional recommended sampling rates will not necessarily result in the national sampling rate.

The results of this analysis have two potential applications.

1. The failure risk can be used to support an assessment of the utility of 100% inspections. Here, we see little risk-based evidence for a reduction in sampling rates.

2. The results can be used to identify those regions that have higher failure risks. However, in this example, there were no failed inspections in any of the regions. The difference in the failure risks between regions reflects that uncertainty that arises because relatively few white rice shipments have been inspected in the two regions with the highest failure risk.

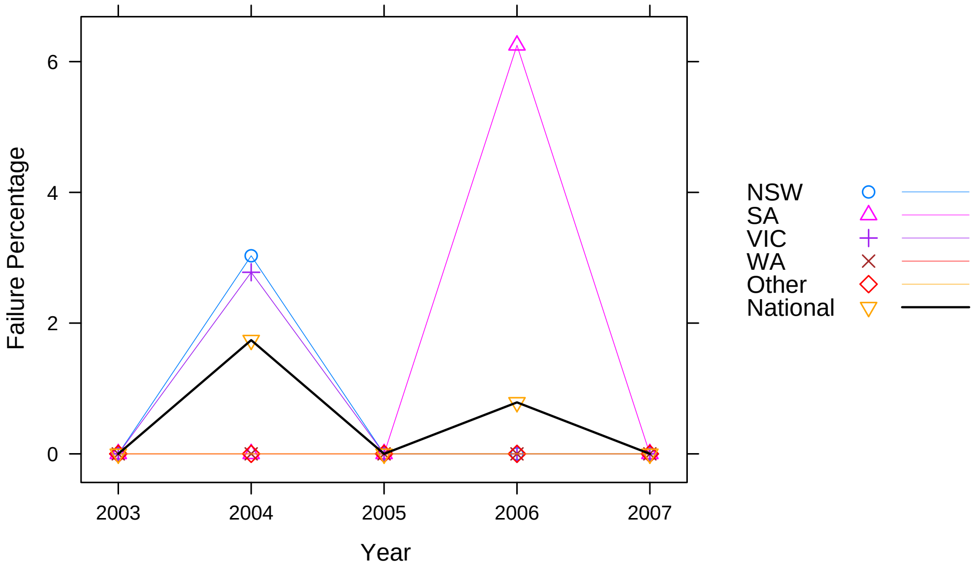

Figure 5.3: Annual failure rate for inspection of white rice, by region and nationally, expressed as a percentage. The thick black line is the national line. *Other* represents NT, QLD, SEQLD, and TAS combined.

5.2.2 Analysis

The IRA conducted by BA included the requirement that polished white rice must be inspected at 100% rate[1]. So far, after numerous years of inspection, only a small number of failures have been found. This observation prompts the question: could some of those resources be diverted to inspecting something else? In order to do so, some sort of case could be mounted to provide feedback on the probability of finding a failure in this commodity to Biosecurity Australia.

We have been provided AIMS data with tariff codes constrained to *rice* commodities. This dataset comprises 71072 observations, detailing 16605 unique AIMS import records across a period of about 5 years. The goods description field comprises 1294 variants, at least some of which are spelling variations. We have records from 837 importers and 1078 suppliers.

In terms of a record of actions taken, there are 22 direction categories and 143 distinct directions. Each AIMS entry contains one or more direction categories. We broke the import records into two groups: those that have at least one direction category that implies a failure, and those records that have no direction categories that imply failure, under ICON condition C8776. Under this tentative rule, any record that has one or more of the direction categories below is defined a failure: *Destruction*, *Heat Treatment*, *Re-Export*, *Other treatments*, and *Failed Food Disposition*.

These records were filtered according to the following steps. The AIMS datasets supplied to us comprised all AIMS records with `RICE` in the goods description record. There were 71072 records. These records were filtered as follows. Records are *included* that:

1. have the word `WHITE` somewhere in the goods description (2637 out of 71072),

[1] The ICON specification for "Rice – White milled or polished" can be found at `http://www.aqis.gov.au/icon32/asp/ex_casecontent.asp?intNodeId=8104383&intCommodityId=7752&Types=none&WhichQuery=Go+to+full+text&intSearch=1&LogSessionID=0`.

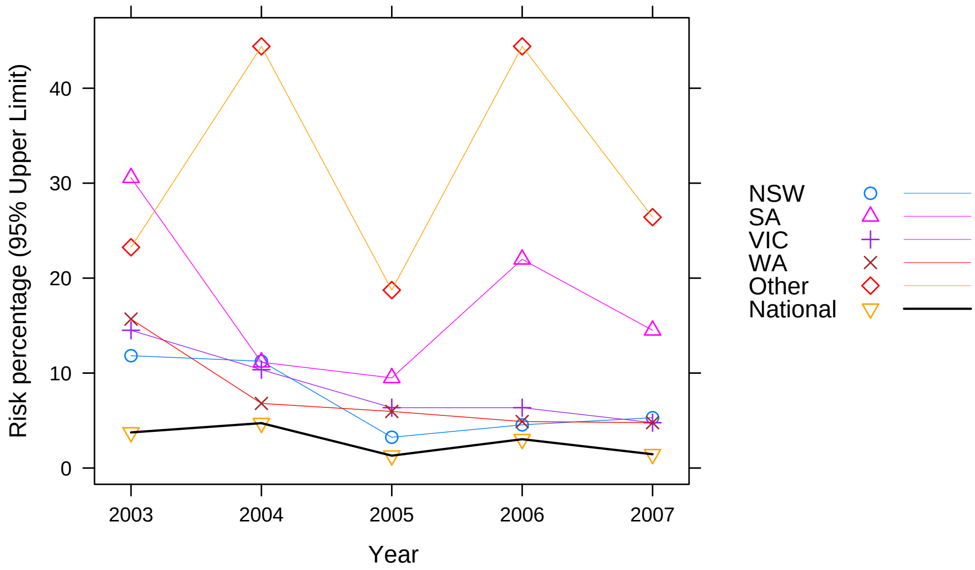

Figure 5.4: Estimated annual 95% failure risk for external inspection of white rice, by region and nationally, expressed as a percentage, for a risk cutoff of 1%. The thick black line is the national line. *Other* represents NT, QLD, SEQLD, and TAS combined.

2. have a direction category that has been defined a failure: *Destruction*, *Heat Treatment*, *Re-Export*, *Other treatments*, or *Failed Food Disposition* (37 out of 2637), and
3. have the word `RICE` (upper or lower case) somewhere in the direction comments or the standard comments (8 out of 37).

These 8 records represent 5 unique consignments. Of these, we judge that records

1. `1S41891361` (`failed condition C8776; paddy grains`)
2. `AAE4APNCC` (`4 ctns white rice required destruction`), and
3. `5F40330187` (`34x1kg Rice destroyed in 'Q' bin by Ross Armstrong 10/3/4 11.15am`)

are likely to be failed consignments of white rice. The first consignment clearly shows evidence of the condition being searched for; the other two records are ambiguous.

A discussion of the nominated analysis strategy, and its implementation in a popular spreadsheet program, is presented in Section 4.5.

To simplify the presentation we chose to construct summary statistics annually, both at the regional and national level. The national-level statistics are the unweighted aggregate of the regional statistics.

Our goal was to obtain a conservative indication of the upper limit of the rate of failure. Therefore we report the upper limit of the one-tailed 95% confidence interval for the estimated failure rate. We can interpret this figure as being a somewhat conservative upper limit for the true rate.

We have converted the rates to percentages for our summary statistics, and for figure 5.4 for ease of interpretation. In all other ways the analytical strategy is identical to that reported in Section 5.1.2.

5.2.3 Further directions

As before, our goal is not to provide the definitive risk analysis for white rice inspection, but rather to use the case study to demonstrate the tools that we suggest will be useful for risk analysis. There is no doubt that a better model could be constructed. For example, if there are sufficient data, then dividing the rice shipments by source country instead of destination might provide a more fine-grained allocation strategy.

The documented analysis treats each year separately, which prevents any sharing of information from year to year. Arguably, the division of collection periods into years is artificial, and a better technique would permit aggregation of data from more than one year. This approach can be easily implemented by accumulating the observed failure and inspection counts across a number of years, possibly in a weighted fashion. The results of a version of this analysis are presented in Section C.2.

5.2.4 What have we learned?

1. AIMS contains all the information necessary for a risk analysis of imported commodities, but some potentially critical information is missing. For example, expanding AQIS information collection to include country of export, overseas supplier, manufacturer, packer, freight forwarder, overseas service provider (eg. fumigator) exporter, carrier, deconsolidator, QAP, stevedore, importer, broker, destination, on-shore fumigator etc., would provide the ability to measure the risk contribution of each component of the import chain, and possibly to take ameliorative action.

2. Similar risk analyses for other commodities can be performed with currently-existing data and tools.

3. Explorations of other ways of profiling white rice might be worthwhile, but with only three failures (or one), not much analytical depth will be possible.

Information Needs

1. AIMS contains all the information necessary for a risk analysis of imported commodities, but as noted above, some potentially useful information is missing.

2. Preparing the data for risk analysis is relatively straightforward but requires tedious hole-plugging and data manipulation,

   (a) Infill missing dates.

   (b) Align descriptors (port names, region names, *etc.*)

   (c) Interpret failure reasons.

3. We have concerns about the reliability of the record of the failure identification; these concerns may be unfounded. Nevertheless, it would be helpful if the exact reason for failure were always included in the AIMS record.

4. The Incidents database is intended to provide comprehensive information on failures, however, record keeping has been regrettably patchy. For example, a number of records in Incidents do not have Quarantine Entry Numbers or dates, and cannot be linked back to the AIMS database. These omissions limit the utility of the entire Incidents database because there is no way to be sure whether the records with missing observations are relevant to an analysis or not.

5.3 AFAS auditing

5.3.1 Summary

The estimated failure rates for AFAS inspections by country and year are reported in Tables 5.3 and 5.4. Some characteristics of the results that stand out are:

- All countries in the analysis are identified by the risk analysis as requiring 100% inspection for auditing purposes.
- Indonesia has improved its estimated risk from about 50% to about 30% between 2006 and 2007.
- No failures were found at 7 facilities in Malaysia.
- India recorded a very high estimated failure rate in 2006.
- Thailand had a higher failure rate in 2007 than did Indonesia, but the greater number of certified entries from Indonesia resulted in a higher risk profile for Indonesia than Thailand in 2007.

The tables provide two candidate approaches to a resource-constrained risk allocation. The first, represented as $\hat{n}_s$ and $\hat{n}_r$ is based on tuning the risk cutoff parameter across the suspect and non-suspect facilities for each country until the total number of facilities to be inspected is equal to a pre-determined cutoff, in this case, 30. The second, represented as $Ratio$, is the normalized product of consignments from that country with the estimated risk value for that country. Resource allocation could proceed relative to the reported value.

We assume for these estimates that inspections are 100% effective.

Table 5.3: Summary of AFAS auditing results for 2007. N_s is the total number of suspect facilities, n_s is the number of those facilities actually inspected, and k_s is the number of suspect facilities that failed. N_n, n_n, and k_n are defined similarly for the non-suspect facilities. $\hat{p}$ is the estimated failure percentage and $\hat{f}_{0.95}$ (%) is the predicted risk, expressed as a percentage. $\hat{n}$ is the tentative future sample size for the random (n) and suspect (s) facilities, constrained to sum to 30 per year (see equation 4.5). *Entries* is the total number of fumigation-certified consignments as reported by AFAS for May 2006-07, and *Ratio* is the relative product of entries and risk (see equation 5.2).

Country	N_s	n_s	k_s	N_n	n_n	k_n	$\hat{p}$	$\hat{f}_{0.95}$ (%)	$\hat{n}_s$	$\hat{n}_n$	Entries	Ratio
Indonesia	22	7	1	65	8	2	22.3	29.9	1.34	0.72	13246	0.547
Malaysia	16	5	0	30	2	0	0.0	14.7	0.89	0.19	9757	0.199
Thailand	11	2	0	27	7	3	30.5	41.3	0.52	2.96	4454	0.254

Table 5.4: Summary of AFAS auditing results for 2006. See Table 5.3 for description of columns.

Country	N_s	n_s	k_s	N_n	n_n	k_n	$\hat{p}$	$\hat{f}_{0.95}$ (%)	$\hat{n}_s$	$\hat{n}_n$	Entries	Ratio
India	25	9	6	106	11	8	71.6	76.3	2.47	0.99	6952	0.430
Indonesia	11	1	1	67	10	4	48.5	53.1	1.00	0.39	13246	0.570

The results of this analysis have four potential applications.

1. The results suggest that, based on our risk analysis approach, 100% sampling is justified for the auditing of all facilities.

2. The results can be used as a reporting statistic for foreign Departments of Agriculture, which are responsible for training and assessing accredited persons, and registering and auditing fumigation companies. For example, based on these results, India could be identified as having a high risk-rate in 2006, and Indonesia could be identified as having a decreasing risk-rate from 2006 to 2007.

3. The results can be used to identify regions of concern for the investment of resource for education, certification, training, and further auditing. For example, the results from 2007 suggest that in 2008, resources should be shifted towards auditing Indonesia, relative to Thailand and Malaysia. Informally, we can include the results from 2006 to add that resources should be shifted towards auditing Indonesia relative to India as well.

4. A future resource-constrained (30 samples) auditing of Indonesia, Malaysia, and Thailand should consider a resource allocation as outlined in table 5.3. This allocation provides the predicted best future balance of risk as we have defined it.

5.3.2 Analysis

The Australian Fumigation Accreditation Scheme (AFAS) audits offshore fumigation facilities in several different countries. When auditing any country it is impossible to visit all the facilities, so only a sample can be audited. AFAS has prior information about facilities, for example, failures detected during inspection of imported commodities. This information is currently used to target AFAS sampling efforts, so that suspect facilities are preferentially sampled. The balance of AFAS inspection efforts is a reasonably random sample of non-suspect facilities, where the sample selection is constrained by location within country.

The above-described approach to data collection is called *stratified sampling*, and predictions of the failure risk facilities can be made by country, and monitored over time. Our prediction of the failure risk is the standard stratified sampling estimate for the failure risk (see *e.g.* Cochran, 1977). The failure risk for a country in a certain year, F, is computed as follows. We assume that before inspection, all facilities in the country can be classified as *suspect* or *non-suspect*. Then the overall failure risk for a country is calculated as:

$$f = \frac{1}{N_n + N_s} \left(N_n \times f_n + N_s \times f_s \right) \tag{5.1}$$

where

f_s is the predicted risk for suspect facilities,
f_n is the predicted risk for non-suspect facilities,
N_s is the total number of suspect facilities, and
N_n is the total number of non-suspect facilities.

Also as an extension of our earlier analysis, to provide more useful guidance than merely "audit everything!", we suggest that the countries identified by AFAS could also be weighted by their relative importance in terms of risk. For this approach we would use the relative product of the import rate of the country and the predicted risk. This index will provide an idea of the relative importance of the risk that each country represents. The index is:

$$R_i = \frac{f_i N_i}{\sum_{j=1}^{c} f_j N_j} \tag{5.2}$$

where we assume that we have c countries for auditing, and that the estimated risk for each country j is f_j, and the total number of consignments that require fumigation certification from country j is N_j. This figure is presented in tables 5.3 and 5.4 as *Ratio*.

We re-emphasize that the results from analyses like these should be treated as a ball-park figure. The practicalities of the physical constraints of auditing will inevitably over-weigh the analysis. However, it is reasonable to believe that, should a trade-off be necessary, Malaysia would be of slightly lower priority than Thailand, and both Malaysia and Thailand would be of lower priority than Indonesia.

It is not presently possible to merge the two resource-constrained allocation appraoches, because we do not have consignment count information broken down by the suspect and non-suspect strata.

5.3.3 Further directions

The approach outlined above provides the essential information for communication and resource allocation. However, other approaches are possible, and other sources of data are available.

For example, the failure rates for the countries could be further weighted by the numbers of consignments that pass through the inspected fumigation facilities. It is reasonable to more heavily weight the failure or non-failure of each facility if the facility processes a larger number of consignments. That information is theoretically available in AIMS, although its extraction may be tedious.

Further information is also available for estimation of the failure rate from the observed percentage of failed fumigation certifications. The certification will identify the AFAS-certified facility from which the failure originated. This information could be used to statistically augment the predictions of the failure risk, using any one of a number of statistical techniques, for example, regression estimation after two-phase sampling. Such augmentation would provide better-quality predictions for the countries and fumigators that are audited, and also the ability to infer failure risks for countries that have not been audited yet.

The stratified sampling approach is simple and robust, but does not make full use of the available data. Other approaches could be deployed that would make similar estimates but also provide the ability to estimate statistics that have bearing on the sampling approach itself.

For example, as a part of the analysis for this case study that is not otherwise reported, we constructed a so-called *generalized linear model* of the failure rate as a function of stratum, country, and year. This model enabled us to conclude that a higher non-compliance identification rate will occur if efforts are focused more towards the suspect stratum; *facilities in the suspect stratum are 1.5 times more likely to be non-compliant compared with the facilities in the non-suspect stratum*. Naturally, a balance should be struck to allow for monitoring and deterrence.

Also, as a part of the analysis for this case study that is not otherwise reported, we constructed a so-called *generalized linear mixed-effects model* of the failure rate as a function of stratum, country, and year. This kind of model allows us to make an estimate called a *shrinkage* estimate, and its justification is as follows. Consider the two estimates for Indonesia: 48.5% in 2006, and 22% in 2007. The difference between the two rates could be explained by success in the AFAS auditing scheme; the facilities are aware that they can be checked and are making improvements. Alternatively, the difference between the two rates could be explained by random selection: with small sample sizes (11 and 15 in each year respectively) it is quite possible that a difference might occur by chance alone. Finally, the difference could be some combination of these two effects. The shrinkage estimate takes account of both possibilities: the difference being due to random chance, and the difference being due to systematic change, and strikes a compromise. A further advantage of a shrinkage estimate is that Malaysia would be assigned a non-zero risk; the estimate of the failure rate would be shrunk away from zero. This estimation technique is also called James-Stein or Empirical Bayes; see, for example, Gelman and Hill (2007) for more details. The advantage of shrinkage estimation is that it provides estimates that are better than traditional estimates from a global point of view, that is, taking account of all the estimation being done.

Finally, the above-recommended prescription focuses on solving the problem based on a very specific interpretation of the problem. We provide an approach that allocates auditing resources according to the estimated risk of each source. Other interpretations are possible, and may indeed be preferable, for example

1. maximizing the number of countries in which AFAS finds at least one non-compliant establishment, or
2. mininizing the expected amount of time that a facility is failing

These different objectives lead to different allocation procedures that have different implications.

5.3.4 What have we learned?

1. AIMS and AFAS jointly hold all the information necessary for a risk analysis to support allocation of auditing resources among source countries.

Information Needs

1. Preparing the data for a simple risk analysis is relatively straightforward.
2. Amassing the data for a more complete risk analysis is a substantial undertaking, requiring the identification of the fumigator certification number corresponding to AIMS records that document fumigation failures.
3. AFAS are aware of the shortcomings of the database available to them and the challenges that arise in connecting it to AIMS. Data in AIMS and Incidents are often inaccurate or missing, owing to
   (a) incorrect data entry procedures,
   (b) required data not being entered,
   (c) useful contextual data not being entered,
   (d) contradictory information in the two databases (*e.g.* insects being recorded as 'dead' in AIMS and 'alive' in Incidents.
   (e) ambiguity in categorical labels.

5.4 Air Cargo (HVLV) surveillance

5.4.1 Summary

Our goal for this case study was to compare the risk between the inspection of HVLV and CAA. Detailed inspection information is available for the former, but not for the latter. Specifically, we lack information about the amount of effort invested in inspection of CAA. Our data suggest that there were 18,085,186 inspectable items in 2007, of which approximately 1150 failed. However, it is not known how many consignments were inspected, and therefore estimating a failure rate and a failure risk is impossible.

We focus on the risk analysis of HVLV. The national failure risk for the for HVLV inspection for 2007 is 0.0196%. The failure rate for HVLV inspection for 2007 is 0.018%. Note that the failure risk is an estimated upper confidence interval for the average rate.

The total number of failures detected nationally from July 2003 until December 2007 is 2966, from a total of 10,586,106 inspections, which is an average rate of 0.028%.

The regional and national predicted annual risks are presented in table 5.5 and figure 5.6. The highest regional failure risk for HVLV inspection for the year of 2007 is 0.22%, in the Far North region. The regional and national failure rates are presented in figure 5.5.

To simplify the explanation, we assume that all possible HVLV have been inspected, and that the inspections are 100% effective.

Table 5.5: Estimated 95% failure risk, average failure rate, and tentative future sampling rate (TFSR) for inspection of HVLV, 2007, presented by region and nationally. The failed and inspected columns refer to the totals for each region in 2007. $\hat{f}_{0.95}$ (%) is the predicted risk, expressed as a percentage. $\hat{p}$ (%) is the average failure rate. TFSR (%) is the tentative future sampling rate, expressed as a percentage, for a risk cutoff of 1% (see equation 4.5). TAS had no inspections in the period.

Region	Failed	Inspected	$\hat{f}_{0.95}$ (%)	$\hat{p}$ (%)	TFSR (%)
Brisbane	140	143778	0.1117	0.0974	0.533
Far North	4	3844	0.2199	0.1041	20.385
NSW	304	1754577	0.0190	0.0173	0.033
NT	2	5906	0.0937	0.0339	10.351
SA	0	38917	0.0049	0.0000	1.384
VIC	16	557516	0.0043	0.0029	0.098
WA	17	151866	0.0164	0.0112	0.370
National	483	2656404	0.0196	0.0182	0.022

Note that the recommended sampling rates at the national and regional levels are not necessarily commensurate. That is, aggregation of the regional recommended sampling rates will not result in the national sampling rate.

The results of this analysis have two potential applications.

1. The reported risk can be used to support an assessment of the utility of 100% inspections. Our results suggest that 100% inspection of HVLV does not necessarily substantially reduce the risk relative to a lesser sampling rate, and that a risk-sensitive approach may recommend sampling HVLV at a lower rate and focusing inspection resources elsewhere, such as CAA.

2. The results can be used to identify those regions that have higher failure risks. Table 5.5 identifies South Australia and Far North as having higher failure risks than the other regions. (See tables C.9 and C.10 for failure and inspection counts by region and year.)

   On the strength of these results, it may be reasonable to maintain the current high inspection rates in these three higher risk regions, but to consider reducing the inspection rates for other regions. In this way the reduction in inspection rates suggested in the previous step can be deployed with best effect.

5.4.2 Analysis

Government legislation specifies that 100% of reportable documents (HVLV, High Volume Low Value, *e.g.* documents) should be examined. When staff have a free moment, they also monitor CAA, opportunistically. CAA seems informally to have a higher rate of failures. The effort to examine CAA is the same as effort to examine HVLV per unit. This is a two-choice investment problem. Limited data are presently available for this exercise[2], the HVLV inspections are well documented but CAA inspections not so.

[2] Ian Jeffries is also interested in an estimate of efficiency of these operations.

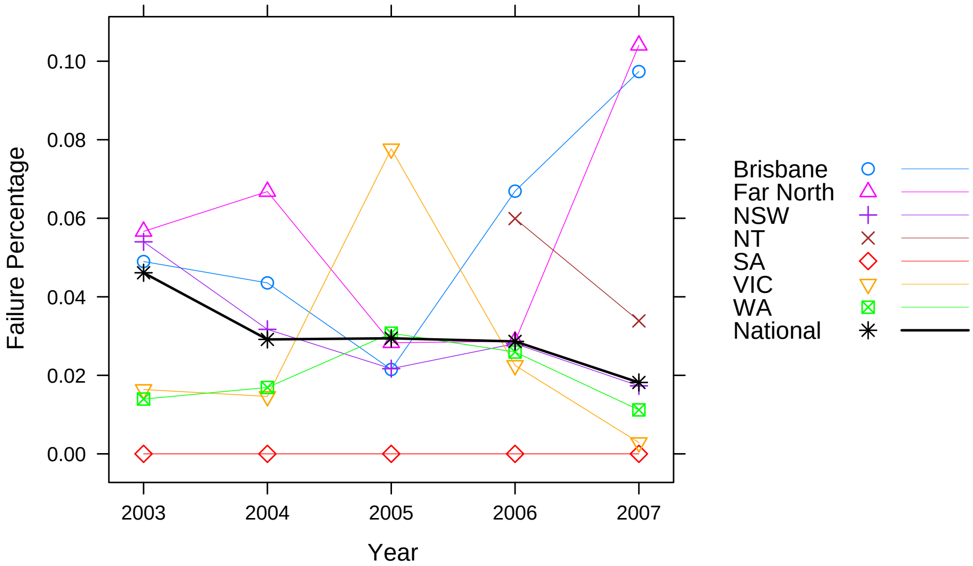

Figure 5.5: Annual failure rate for inspection of HVLV, by region and nationally, expressed as a percentage. The thick black line is the national line. NT had negligible inspections before 2006.

The HVLV inspection data that we received from AQIS were monthly counts of inspections and failures, by region. To simplify the presentation we chose to construct summary statistics annually, both at the regional and national level. The national-level statistics are the unweighted aggregate of the regional statistics.

Our goal was to obtain a conservative indication of failure risk. Therefore we report the upper limit of the one-tailed 95% confidence interval for the estimated failure rate. We can interpret this figure as being a conservative upper limit for the true rate. Otherwise the analysis was performed as laid out in Section 4.5.

We converted the rates to percentages for our summary statistics, and for figure 5.6 for ease of interpretation.

5.4.3 Further directions

Our goal is not to provide the definitive risk analysis for HVLV inspection, but rather to use the HVLV case study to demonstrate the tools that we suggest will be useful for risk analysis. Given sufficient information there is no doubt that a better allocation could be constructed.

Dividing the shipments by destination is an artificial approach that we adopted because the data were readily available. It would make more sense to divide the shipments by source country, if that information were available.

The documented analysis treats each year separately, which prevents any sharing of information from year to year. Arguably, the division of collection periods into years is artificial, and a better technique would permit aggregation of data from more than one year, which would simplify the identification of seasonal differences within the year. This approach can be easily implemented by accumulating the observed failure and inspection counts across a number of years, possibly in a weighted fashion. The results of a version of this analysis are presented in Section C.3.

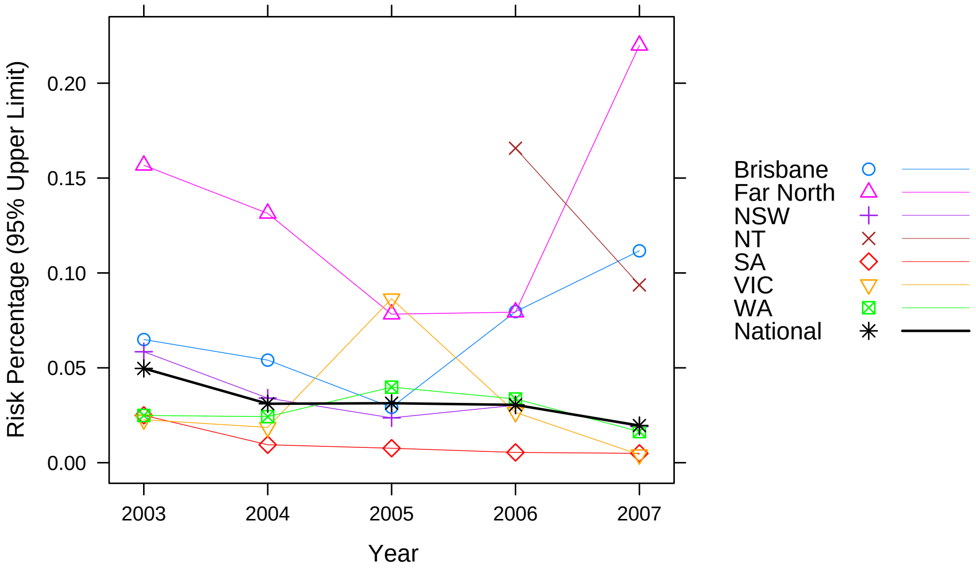

Figure 5.6: Estimated annual 95% failure risk for inspection of HVLV, by region and nationally, expressed as a percentage, for a risk cutoff of 1%. The thick black line is the national line. NT had negligible inspections before 2006.

5.4.4 What have we learned?

1. AQIS has the necessary data to question whether 100% inspection of HVLV is worthwhile from the point of view of risk analysis.

Information Needs

1. AQIS already aggregates inspection data for HVLV inspection by month and year, and by the region to which the HVLV is delivered. This dataset was used for the analyses reported above.
2. A more precise profiling approach would be possible if more information about the individual items were readily available. For example, failure rates might vary by source country, and profiles that were constructed with this knowledge would enable better-targeted surveillance.
3. At present AQIS doesn't collect data that would be necessary for a risk analysis of CAA.

5.5 ICE

5.5.1 Summary

Substantial progress towards resolution of this case study will be possible when the ICE records are connected with their corresponding AIMS records.

5.5.2 Analysis

Import Clearance Effectiveness (ICE) audit selections are presently a mixture of haphazard selection by an officer and automatic selection by ICS. A plausible improvement is in provision of risk-based tools to focus greater attention upon those consignments that are more likely to fail. We would need to fit a model to identify and predict systematic differences in the rate of occurrence of failures.

The ICE data comprise:

1. 33354 observations,
2. 19823 AIMS entry numbers,
3. two inspection locations,
4. eight city locations (internal)
5. 18914 goods descriptions, and
6. > 25 failure flags.

We refer to the imported food case study (Section 5.6) for a demonstration of the proposed methodology, although we would not apply the rescaling step that is used to constrain the overall sample size for ICE data.

5.5.3 What have we learned?

1. AIMS and ICE jointly hold all the information necessary for a risk analysis to support allocation of auditing resources among different consignment profiles.
2. These data are not in a readily analyzable form, owing to the records being split across the ICE and AIMS databases, and the lack of a means of easily connecting the two.

Information Needs

We were unable to explore the full potential of this case study, owing to the lack of available data.

1. Preparing the data for a simple risk analysis is relatively difficult. The ICE database does include AIMS record numbers, but there does not seem to be a straightforward way to automatically obtain the AIMS records corresponding to those numbers, except by intervening analytical steps that may require computing resources beyond those currently available. The major challenge is that AIMS was built to store data, and enabling the retrieval of data has not been prioritized.

   It is possible to obtain flat-file dumps of AIMS data; this is the basis upon which we analyzed the white rice case study. It would theoretically be possible to dump the entire AIMS dataset into a single, searchable file, and then locate the ICE records within it.
2. Managers of the ICE data are aware of the shortcomings of their database and the challenges that arise in connecting it to AIMS.

5.6 Imported Food

5.6.1 Summary

If Imported Food (IF) acquires the ability to profile their approaching consignments, and the profiles are used to selectively alter their sampling rates, subject to an overall constraint, then the detection rate of failures may increase. Table 5.6 (p. 50) summarizes the approach.

We found that by applying a risk-based profile that uses the identity of the proposed test as a category (see table 5.6), the rate at which failures are detected increases, without adding to inspections. Using independent data we estimate that the detection rate could be more than doubled. For example, using the nominal sampling approach led to the detection of 770 non-conformities in the first six months of 2007. We estimate that, using one of the risk-based approaches described below, the detections could increase to close to 2000 (last row, table 5.6).

As an example of the application of the profile, the test for mercury should be applied to 16% of those consignments that are identified as requiring the test for mercury (see table 5.6). Currently, by default, the test is applied to 5% of the consignments. On the other hand, visual tests are so unlikely to detect failures that they should only be applied to 0.3% of the consignments that are identified as requiring a visual test.

Note that the specific approach that we suggest here, as an example, might not be practical. In order to use this approach, it would be necessary to nominate a test for *every* consignment, nominate a sampling rate based on the identify of the test, and then determine whether the consignment should be inspected. If the nomination of a test is relatively easy to automate then the process can be automated. Also, the approach could be further extended by possibly separating the tests into classes such as those that require lab work and those that do not. Automated allocation might be plausible for one class but not the other.

Some consignments might have more than one nominated test, in which case we might apply a unique determination of whether a test should be executed for each test instead of each consignment. That might also allow for a conditional testing approach, if failures tend to be correlated within consignments.

5.6.2 Analysis

Imported food is legislated by FSANZ[3] to be inspected at one of a number of rates: 5% if not categorized as requiring active surveillance, 10% if categorized as requiring active surveillance, and 100% inspection if categorized as risky. *As far as we are aware*, the level of aggregation to which this rate applies is not stated. For example, does it mean 5% of all imported commodities, or exactly 5% of every commodity type? If the former, then it is possible to be flexible in terms of what is sampled and when it is sampled, whilst maintaining a rigorous 5% level.

However, in order to benefit from such an approach it would be necessary for AQIS/IF to have control over the random selection of items for inspection, which is currently performed in ICS. Here we use IF monitoring data which will enable us to estimate the potential benefits from such a program were it possible to be deployed.

The IF data comprises several comparatively fine-grained categorical descriptors that could be candidates for profile construction, for example:

1. 36195 observations
2. 7428 Quarantine entry numbers
3. 441 unique tariff numbers
4. 107 countries
5. 6194 goods descriptions
6. 42 testnames

Using the previously-developed approach for risk profiling becomes problematic as the number of categories increases, and the categories become emptier, due to the problem of scale outlined in Section 4.5.4. To simplify this analysis we assumed that each observation represented a unique consignment in the fitting and the test data.

We constructed four distinct model-based approaches to estimating the risk, as well as the estimate outlined in Section 4.5.

[3] www.foodstandards.gov.au/

Model construction

We fitted a generalized linear model (glm) to estimate the risk profiles at a finer scale. Continuing from the model expressed in equation 4.1, we index each commodity j with group membership i, where i could be for example the supplier, or the test based on a profile, so

$$k_{ij} \sim \text{Binomial}(p_{ij}, n_{ij}), \tag{5.3}$$

where k_{ij} denotes the observed number of failures from commodity j in group i, n_{ij} is the number of consignments inspected from commodity j in group i, and p_{ij} is the unknown probability of failure. Then we construct a glm for p_{ij} as follows:

$$\ln\left(\frac{p_{ij}}{1-p_{ij}}\right) = \gamma_i + \epsilon_{ij} \tag{5.4}$$

where γ_i is an unknown group-specific parameter and $\epsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$ is a commodity-specific error (also called a residual). We note in passing that there are many other approaches to the same problem, such as generalized linear mixed models (glmm, see below), support vector machines, and random forest models (see, *e.g.*, Hastie et al., 2001). A broad-reaching simulation study would be very useful to compare these options for AQIS operations.

Two of the alternative proposed risk estimates for group i are then based on the estimated non-conformity *rate* (see equation 5.5), and the estimated non-conformity *risk* (see equation 5.6).

$$\hat{f}_{pi} = \frac{\exp(\hat{\gamma}_i)}{1+\exp(\hat{\gamma}_i)} \tag{5.5}$$

$$\hat{f}_{ri} = \frac{\exp(\hat{\gamma}_i + 1.96 \times \hat{\sigma}_{\gamma_i})}{1+\exp(\hat{\gamma}_i + 1.96 \times \hat{\sigma}_{\gamma_i})} \tag{5.6}$$

The final step is to scale these risk values linearly until the nominated sampling rate is matched ($\pi = 0.05$ in this case). Thus, in each case,

$$\text{TFSR}_i = \min(\delta \times f_i, 1) \tag{5.7}$$

where we solve

$$\pi = \frac{\sum\sum \min(\delta \times f_i, 1) \times n_{ij}}{\sum\sum n_{ij}} \tag{5.8}$$

The TFSR_i are the values reported in table 5.6.

We also deployed a generalized linear mixed-effects model (glmm). Such models use shrinkage estimation, which is known to provide better-quality estimates in cases of sparse data (see Section 5.3.3). We construct a glmm for p_{ij} as follows. Firstly we extend equation 5.4:

$$\ln\left(\frac{p_{ij}}{1-p_{ij}}\right) = \mu + \gamma_i + \epsilon_{ij} \tag{5.9}$$

where μ is the underlying failure rate across all commodities, $\gamma_i \sim \mathcal{N}(0, \sigma^2_{\gamma_i})$ is a (shrunk) group-specific correction, and $\epsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$ is a commodity-specific residual. Note that the distributional assumption on γ_i induces the shrinkage.

The final two proposed alternative risk estimates for group i are then based on the estimated non-conformity *rate* (see equation 5.10), and the estimated non-conformity *risk* (see equation 5.11).

$$\hat{f}^m_{pi} = \frac{\exp(\hat{\mu} + \hat{\gamma}_i)}{1+\exp(\hat{\mu} + \hat{\gamma}_i)} \tag{5.10}$$

$$\hat{f}^m_{ri} = \frac{\exp\Big(\hat{\mu} + \hat{\gamma}_i + 1.96 \times \sqrt{\hat{\sigma}^2_{\mu} + \hat{\sigma}^2_{\gamma_i}}\Big)}{1 + \exp\Big(\hat{\mu} + \hat{\gamma}_i + 1.96 \times \sqrt{\hat{\sigma}^2_{\mu} + \hat{\sigma}^2_{\gamma_i}}\Big)} \tag{5.11}$$

As above, the final step is to scale these risk estimates linearly until the nominated sampling rate ($\pi = 0.05$) is matched.

Increasingly fine-grained risk estimates can be constructed by adding further corrections. For example, in the current case study we might try a profiling model that uses test- and country-based corrections.

Model testing

IF provided us with two datasets, comprising inspection information from the second half of 1996 and the first half of 1997. We used the first dataset, hereafter called the *fit* data, to estimate TFSR_i, and the second dataset, the *test* data, to compare them. Our test was as follows.

1. Compute risk estimates from the fit data using method outlined above.
2. Sum the detected failures. If we have c categories, then

$$E(k) = \sum_{i=i}^{c} \times k_i \tag{5.12}$$

3. Compute the expected number of detected failures for the test data based on a sample where the actual sampling rate of each individual commodity was based on whichever of the $\hat{f}_i$ was being used, but the overall rate is constrained to be 5%.

$$E(k) = \sum_{i=i}^{c} \frac{\hat{f}_i}{0.05} \times k_i \tag{5.13}$$

A comparison of the results of steps 2 and 3 demonstrates the improvement in the rate of detection of failures that arises from profiling whilst maintaining the 5% inspection rate. See the bottom row of table 5.6 for an example of risk profile-based sample allocation.

5.6.3 Further directions

Risk analysis is applied in concert with data mining using R in other organizations, for example the Australian Taxation Office (ATO Analytics section; 16 data mining specialists supporting 120 analysts, providing an over-arching framework for risk management[4]). Existing data mining techniques could be deployed on AQIS data resources, such as AIMS data, to construct risk profiles that may have better properties. Hastie et al. (2001) provide an excellent starting point.

Also, the reported analysis deliberately fails to distinguish between sample-based variation and the variation of the underlying process. That is, there may be uncertainty about future failure rates because (a) there is uncertainty about the true future rate, given that the current rate is assumed to be known, and (b) there may be uncertainty about the current rate as well as the future rate. Given enough information it would be possible to estimate the contribution of each source of variation, which may allow for the construction of better risk estimates. That is, we could use the observed year-to-year variation to provide a less conservative risk analysis approach; presently we conflate the year-to-year variation with the within-year variation.

[4] Graham Williams, Rattle: Data Mining with R, ASC Workshop 28/6/2008.

5.6.4 What have we learned?

1. Careful application of risk-based profiles can increase the detection rate of failures without increasing the overall inspection load.

2. Model construction for simple cases is relatively straightforward.

Information Needs

The imported food case study demonstrated the utility of a readily-available, relatively complete database. Further gains still could have been made, and further ideas tested, if

1. we had access to data collected in previous years - this exercise was performed on only 12 months of data;

2. the full imported food approach dataset were available, as it comes to the ICS.

Table 5.6: Summary of diagnostic-test-based profile model for imported food data. $\hat{p}_p$ and $\hat{p}_v$ are the estimated failure rates for the test and validation data. The other columns are recommended sampling rates. $\hat{f}_\beta$ is from Section 4.5, $\hat{f}_p$ and $\hat{f}_r$ are the rate- and risk-based estimates from the glm approach (equations 5.5 and 5.6), and $\hat{f}_p^m$ and $\hat{f}_r^m$ are the rate- and risk-based estimates from the glmm approach (equations 5.10 and 5.11). The last row reports the estimated number of detections. The data used to compute these predictions are presented in table C.13.

Test	$\hat{p}_p$ (%)	$\hat{f}_\beta$ (%)	$\hat{f}_p$ (%)	$\hat{f}_r$ (%)	$\hat{f}_p^m$ (%)	$\hat{f}_r^m$ (%)	$\hat{p}_v$ (%)
1 3 Dichloropropanol	0.0	11.7	0.0	0.0	0.9	3.7	0.0
3Monochloropropan1,2Dl	1.1	14.3	2.1	8.6	2.4	5.7	0.0
Aflatoxin	3.5	5.4	6.4	7.9	6.4	7.0	3.9
Artificial Sweeteners	5.3	100.0	9.6	34.7	7.1	19.2	0.0
B Cereus	17.9	100.0	32.9	45.3	29.3	33.4	1.5
Bse Cert Check	0.8	1.2	1.5	2.6	1.6	2.2	0.3
Cadmium	5.1	46.6	9.4	10.8	9.3	9.8	4.5
Carbadox	0.0	100.0	0.0	0.0	1.4	6.5	0.0
CheckLablFrAddtv(nTst)	30.0	100.0	54.9	85.5	38.3	52.1	0.0
Chloramphenicol	0.0	7.3	0.0	0.0	0.7	2.6	0.0
Coliforms	0.0	100.0	0.0	0.0	3.0	21.7	0.0
Colour Screen	2.9	16.7	5.2	10.7	5.0	8.1	1.1
Commercial Sterility	3.4	100.0	6.2	8.9	6.1	7.4	0.0
Composition	42.2	100.0	77.3	72.4	76.5	60.0	34.1
Domoic Acid	0.0	14.1	0.0	0.0	1.1	4.5	0.0
E Coli	5.2	34.1	9.5	9.7	9.5	9.4	2.4
Ethylene Chlorohydrin	0.7	2.6	1.3	3.3	1.4	2.6	0.3
Histamines	1.8	1.2	3.2	4.0	3.2	3.6	2.2
Inorganic Arsenic	100.0	100.0	100.0	100.0	30.8	74.6	100.0
Labelling	5.2	2.4	9.5	8.2	9.5	9.0	5.8
Lead	7.1	100.0	13.1	21.2	12.0	16.0	0.0
Listeria Monocytogenes	1.3	1.5	2.4	3.2	2.5	2.8	1.5
Malachite Green	0.0	21.2	0.0	0.0	1.5	7.2	0.0
Mercury	2.9	5.1	5.2	6.4	5.2	5.8	2.5
Nitrofurans	7.0	100.0	12.8	16.3	12.4	13.8	5.6
Officer To Assign Test	0.0	100.0	0.0	0.0	2.6	16.8	0.0
OystersExJapan/Korea	100.0	100.0	100.0	100.0	30.8	74.6	100.0
Patulin	0.0	100.0	0.0	0.0	2.4	15.2	0.0
Pesticides	0.6	0.4	1.1	1.4	1.1	1.3	0.3
Ph	0.0	100.0	0.0	0.0	0.3	1.0	0.0
Psp Toxin	0.0	14.1	0.0	0.0	1.1	4.5	0.0
Radiation	100.0	100.0	100.0	100.0	30.8	74.6	0.0
Salmonella	1.1	0.6	2.0	2.3	2.0	2.2	0.8
Standard Plate Count	2.7	6.4	5.0	7.2	4.9	6.1	4.6
Staph Enterotoxin	0.1	100.0	0.2	0.9	0.4	0.8	0.0
Staphylococci	0.8	4.0	1.5	3.8	1.7	3.0	0.0
Streptomycin	0.0	100.0	0.0	0.0	3.0	21.7	0.0
Sulphonamides	0.0	100.0	0.0	0.0	3.0	21.7	0.0
Sulphur Dioxide	2.7	13.9	5.0	9.1	4.9	7.2	1.0
Tetracyclines	0.0	100.0	0.0	0.0	2.9	20.2	0.0
Vibrio Cholerae	1.1	6.9	2.0	5.0	2.1	3.9	0.0
Visual	0.1	0.1	0.2	0.3	0.2	0.3	0.2
OVERALL	935	1819	2026	1879	1995	1817	770

6

Recommendations

Much can be done with existing data and tools, and very much more may be possible.

6.1 Following this report

1. *We recommend the establishment of an internal risk team.*

   The ACERA team talked with national and regional AQIS officers for only a few days, and in so doing developed a substantial list of information needs. This report documents strategies for the resolution of only a handful of the challenges. It is very clear that considerable scope exists for a continuation of these efforts.

   We suggest that the brief of such a team would be to identify and resolve the many risk-related challenges that AQIS faces now and in the future. Numerous examples of potential studies are documented in this report, a number of which were identified through the ACERA visits to regional offices. We advocate that such a team should undertake regular communication with regional offices in order to identify new opportunities, and to provide regional officers with timely and appropriate feedback for existing concerns.

   A precedent for such a resource can be found in the Australian Taxation Office (ATO Analytics section; 16 data mining specialists supporting 120 analysts, providing an over-arching framework for risk management[1]).

2. *We recommend the development of a collection of easily-grasped decision scenarios, which may be based on or related to the case studies described in this report, to allow AQIS officers to formally calibrate the risk-based approaches.*

   A common response from AQIS officers to presentations of this report has been that the nominated risk-analysis approach is more conservative than necessary. Balancing comfort with resources is a very important consideration. This calibration would involve computing the inspection prescriptions given the risk parameters, and would allow risk parameters to be selected that reflect an underlying comfort with the risk-level/resource-use trade-off. The resolution of this trade-off in simple scenarios could then be used in more complicated situations, for which intuition is not so readily available or reliable.

3. *We recommend that AQIS IC undertake a simulation study to compare the properties and performance of the different risk-estimation approaches that have been canvassed in this report, including data mining.*

   We have proposed a number of different risk-analysis approaches in this report. It would be very useful to formally compare these approaches with one another, also taking account of some

[1] Graham Williams, Rattle: Data Mining with R, ASC Workshop 28/6/2008.

of the choices that are required. This could be accomplished using simulations and existing datasets.

Also, more sophisticated risk analyses are possible for some of the case studies. Such analyses might involve predictions across time, possibly taking account of seasonal patterns, supplier, or region of origin. Formally testing such analyses using simulations would allow for better choices of the level of detail that is needed for risk management.

4. *We recommend that a Multi-Criteria Decision Analysis implementation program include tests of alternative value trees, weights and aggregation methods, with the ultimate aim of finalizing a tool that could be used in risk minimization and feedback systems.*

   The risk analysis framework that we implemented for this project requires further testing and development before it will be operational.

6.2 What is possible now

Here we describe the risk-based developments that are possible with current data and readily-available analytical tools.

5. *We recommend that AQIS IC undertake risk analyses of the kinds demonstrated for aircans (Section 5.1), white rice (Section 5.2), and HVLV (Section 5.4) for other commodities or processes of interest.* Examples of such risk analyses are:

   (a) *We recommend that AQIS IC undertake a risk analysis for ECIR.*
   The external container inspection is an expensive undertaking. AQIS IC should use the tools developed in this report to assess whether or not the expense is justified.

   (b) *We recommend that AQIS IC undertake a risk analysis for gun handle inspection.*
   Presently gun handles are inspected at 100%, despite a very low or zero failure rate. Guns are typically imported from countries that have low-risk wood products, for example the USA and Germany. Gun handles are usually constructed to high standards with high-quality timber or plastic. The data are not currently organized but should be accessible through AIMS, as firearms have a unique tariff. AQIS IC should use the tools developed in this report to assess whether or not the expense is justified.

   (c) *We recommend that AQIS IC undertake a risk analysis for QAP audits.*
   Quarantine Approved Premises (QAP) nominally require two audits per year. The Brisbane office has been trialling a system whereby field officers are given a checklist during their routine visits, approximately monthly. These checklists are used as a basis for prioritizing audits. The checklists act as a rolling low-intensity audit, but can also be used as a motivation and guide for helping QAP managers to remain compliant, or regain compliance. Such a strategy reflects the AQIS thinking that intervention is better than suspension. This is a two-phase sampling problem. The data are available, but the initiative is young, so the data are comparatively few. AQIS is also consolidating its audit strategy for compliance agreements and QAP audits.

   (d) *We recommend that AQIS IC undertake a risk analysis for fresh produce inspection.*
   Fresh produce inspection involves the sampling of produce to a predetermined level, for example, inspection of 600 units. Once a failure has been detected in fresh produce inspection, the consignment will be sent for treatment, for example, fumigation. Field sampling crews are reluctant to continue sampling under these circumstances, considering it a waste of time and resources. That is, *once a failure is found*, why continue to sample?
   A two-phase sampling rule could be developed for certain products. Two-phase sampling involves the deployment of two sampling schemes, one of which is expensive and painstaking, the other of which is cheaper and faster, and statistical calibration is used to ensure

that high data quality is maintained. As an example of such a design, only every second sample could be completed according to the current specifications, but the interleaving samples are run as follows: sample until

1) the current specifications are satisfied, *or*
2) a non-conformity is detected, in which case
   A. stop sampling,
   B. report the non-conformity and the number of items inspected, and
   C. send the consignment in for fumigation.

This strategy would marginally reduce the amount of product data available to OSP but would substantially reduce the field workload.

However, adoption of such an approach does entail a loss of information. While stopping inspection as soon as an indication that that fumigation is required will minimise inspection effort, less information is obtained about the level of infestation, and more importantly, on whether there are more than one different types of non-compliance. The importance of this extra information should be balanced against the effort required to collect it.

(e) *We recommend that AQIS Imported Food undertake a risk analysis for the potential utility of profiling for risk management.*

The IF case study documented in Section 5.6 demonstrates that profiling increases the chance of finding failures. Here we outline a further development that uses the developed profiles to inform whether or not shipments should be released before the test results are known.

Imported food is legislated by FSANZ[2] to be inspected at one of a number of rates: 5% if not categorized as requiring active surveillance, 10% if categorized as requiring active surveillance, and 100% inspection if categorized as risky. The protocol for action after sampling also differs by classification. For example, the random samples of commodities not requiring active surveillance involve test and release, with follow-up to relevant parties if a problem is detected. That is, a sample is taken and the product is released, so the test results are not usually known in time to intercept the product. On the other hand, risky imports are under a test and hold system, which means that the product is held until it is cleared. Any randomly-sampled products that show systematic non-conformities will be released under the present system. A candidate risk-aware solution would be to use a statistical model to predict non-conformity rates in different commodities, regions, and suppliers, for example, so that systematic failures could lead to an escalation from test and release to test and hold. This approach could reduce the level of risk that results from the current protocol, and is similar to the Pathway Node Register.

6.3 What is possible with modest internal change

Here we describe the developments that would be possible with modest change, which we define as change within the purview of AQIS.

6.3.1 More complete data collection and storage

6. *We recommend that AQIS IC collect and store data about the identity of each step of the import process.*

   Expanding AQIS information collection to include such information as country of export, overseas supplier, manufacturer, packer, freight forwarder, overseas service provider (eg. fumigator)

[2] http://www.foodstandards.gov.au/

exporter, carrier, deconsolidator, QAP, stevedore, importer, broker, destination, on-shore fumigator etc., would provide the ability to measure the risk contribution of each component of the import process, and possibly to take ameliorative action.

7. *We recommend that AQIS IC collect and store data on conformities as vigilantly as it does on non-conformities.*

   Information on conformities is essential as it provides the context for the analysis of data about non-conformities. Without knowing the background approach rate there is no way to identify the risk of failure.

8. *We recommend that AQIS IC develop and implement flexible data collection, storage, and access protocols to enable risk analyses for a wider range of inspection challenges.*

   During the course of our meetings with AQIS officials we uncovered numerous opportunities for process improvement with AQIS that required specialized, snapshot data collections. Examples of such risk analysis opportunities are:

   (a) *We recommend that AQIS IC undertake a risk analysis for Personal Effects (PE) inspection.*
   PE inspections always involve sampling. PE could be objectively profiled for greater or lesser threat status if specific data on failures and successes, including kind and location, were available. Other information could also provide profiling data, such as: country of origin, whether or not professionally packed, relative amount of detail provided on manifest, and some record of the non-conformity history and type. Such an approach requires more detailed data capture than is presently performed, for example detailed information on the characteristics of consignments that showed no non-compliance.

   (b) *We recommend that AQIS IC undertake a risk analysis for the quantity of bark on packaging material.*
   ISPM 15 regulations state that there shall be no bark on any wood used for packing materials. However, the element of real concern is not bark itself but rather the insects and diseases that it may house. This reflection suggests the question: how much bark is too much? It may well be that inspection for minute amounts of bark is unwarranted; bark that is sufficiently small to require close inspection may be too small to pose any significant threat - and it may not. Potential covariates include country of origin, commodity, supplier, size of bark pieces, contiguity of bark pieces, time of year, species of concern. In theory this study could be resolved with existing data, but for the concerns with matching AIMS and Incidents data already noted.

   (c) *We recommend that AQIS IC undertake a risk analysis for seed contamination of new cars.*
   Recently new cars from Asian countries have been harbouring seeds from plants of interest. The seeds are small and hard to find, and often nestle in hard-to-inspect locations of the vehicles, such as in bonnet rims and underneath the chassis. The current IC position is surveillance; there is no profile against the commodity, although inspections are being accommodated by the importer. The sampling protocol is as follows: the first ten or so cars are examined very closely, after which the severity of inspection decreases. If more than one or two seeds are found, then the entire batch of vehicles is ordered in for cleaning and inspection, including inspecting the underside by hoisting the car using a light crane. The latter operation, which is mandatory, represents a considerable bottleneck for the QAP.

   If a detailed data collection exercise were in place, then it might be possible to streamline the operation from a number of points of view. For AQIS IC, the goal should be to inspect as few cars as possible before making the decision to send the batch in. This is a sample design problem that requires information on seed counts and location by car type and other variables, to target sampling operations. For the QAP, the goal should be to have as few cars on the hoist as possible, because that is the processing bottleneck. Thus the QAP may find it worthwhile to clean all cars before inspection in certain seasons, or to try to argue that a batch is not necessarily a logical unit for surveillance decision-making, or to install a ramp to facilitate undercarriage inspections. In any case, both partners would likely benefit from a more rigorous data collection and analysis approach.

6.3.2 Streamlined data access

9. *We recommend that AQIS IC research, acquire, and deploy data manipulation tools that will enable the extraction, manipulation, and analysis of the data that are stored but inaccessible in AQIS databases such as AIMS.*

   The white rice case study was undertaken using R, a free, open-source statistical environment, upon a flat-file dump from the AIMS database. The case study demonstrated that such an analysis is possible, but it was not simple or convenient. In order to benefit most fully from risk-based management, it is necessary for AQIS IC that their data be readily available for analysis with as few impediments as possible.

   In many cases, the necessary information for risk analysis is probably available in AIMS and Incidents, but the information gaps in each are problematic. For example, it is useful to know why a commodity failed: it may have been due to a condition of interest, or it may be due to non-commodity failure. The reasons for commodity failure are recorded in AIMS but not using consistent vocabulary or reporting standards, and the Incidents database lacks AIMS identifiers for many of its records.

   As a further example, it should be possible to easily locate and extract the full AIMS records for the ICE data. This development would enable analysts to more easily create and test a wider range of risk analyses for commodities or activities of interest, and would greatly ease automation of this process.

   In the interim, it might be possible to use an intermediate step to simplify data manipulation. For example, coarse dumps from AIMS might be subsequently refined as we have done, using R, or using a database tool such as PostgreSQL or MySQL.

6.4 What might be possible

Here we describe the developments that would be possible with external change, by which we mean change within the environment within which AQIS operates.

6.4.1 Ownership of Data

10. *We recommend that AQIS should get joint ownership of the incoming data from ICS.*

    AQIS receives data from ICS about only those commodities that match the AQIS filters. AQIS receives no data about the commodities that do not match its filters. As we have noted above, collecting information about conformities is key to a risk management approach. Collecting information only on non-conformities does not suffice.

    AQIS should somehow get ownership of the incoming data in order to

    (a) know the approach rate, and
    (b) flexibly and promptly apply relevant filters.

Without ownership of the incoming data, the full implementation of a risk management approach to Import Clearance is compromised.

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.

ANAO (2005). Managing for quarantine effectiveness - Follow-up: Department of Agriculture, Fisheries and Forestry, Biosecurity Australia. Audit report no. 19, 2005-06 performance audit, Australian National Audit Office, Commonwealth of Australia, Canberra.

Brown, L. D., Cai, T. T., and DasGupta, A. (2001). Interval estimation for a binomial proportion (with discussion). *Statistical Science*, 16:101–133.

Cai, T. T. (2005). One-sided confidence intervals in discrete distributions. *Journal of Statistical Planning and Inference*, 131:63–88.

Cochran, W. G. (1977). *Sampling techniques*. John Wiley & Sons, Inc., 3rd edition.

Gelman, A. and Hill, J. (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press. 625 p.

Hastie, T., Tibshirani, R., and Friedman, J. (2001). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer.

Krishnamoorthy, K. (2006). *Handbook of Statistical Distributions with Applications*. CRC Press. 346 p.

Madden, L. V., Hughes, G., and Munkvold, G. P. (1996). Plant disease incidence: inverse sampling, sequential sampling, and confidence intervals when the observed mean incidence is zero. *Crop Protection*, 15(7):621–632.

R Development Core Team (2008). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

Appendix A

Glossary of Acronyms

ACS Australian Customs Service
ACU Air Cargo Unit
AEP Automatic Entry Program
AFAS Australian Fumigation Accreditation Scheme
AIMS AQIS Import Management System (Nat. Admin. Robert Savage). State databases are held locally, and each one is also accessible remotely.
AQIS Australian Quarantine and Inspection Service
BA Biosecurity Australia
BAS Broker Accreditation Scheme, verified by (minimum) 3% audit of documents.
CA Compliance Agreement
CAA Cargo Assurance Air - also called *freeline*.
CR Cargo Report
ECIR External Container Inspection Regime
FC Front Counter
FCL Full Container Load
FID Full Import Declaration, often lodged electronically by broker, > $1000 au
GAS Giant African Snail
HVLV High volume low value (*e.g.* documents through couriers)
ICE Import Clearance Effectiveness
ICON Import CONditions (online database of conditions for import) URL is `http://www.aqis.gov.au/icon32/asp/ex_querycontent.asp`
ICS Integrated Cargo System, with database held centrally in Canberra
IF Imported Food
IPPC International Plant Protection Convention, publishes guidelines for *e.g.* sampling. `https://www.ippc.int/IPP/En/default.jsp`
IPU Industry Partnerships Unit = QAP + CoReg
IRA Import Risk Analysis
ISPM 15 International Standards for Phytosanitary Measures No. 15: Guidelines for regulating wood packaging material in international trade
ISTA International Seed Testing Association
LCL Less than a full Container Load
OSP Operational Science Program
QAP Quarantine Approved Premises
QICD Quarantine Insect Collection Database
QPR Quarantine Premises Register
SAC Self-Assessed Clearance, electronic lodging < $1000 (AU)
UPE Unaccompanied Personal Effects

Appendix B

Alternative Risk Methods

This appendix presents some other risk analysis approaches that were considered to be less useful.

B.1 Introduction

Large numbers of objects (containers, consignments, lots) from many sources (countries) pass through an inspection point. We define

p_x is the probability an object from source x contains a threat, T,

N_x is the number of objects from x that pass through an inspection point in some unit time interval,

n_x is the number of objects from x inspected in the unit time interval for the common threat, T.

Economic and other constraints limit the total number of inspections, I, to

$$\sum n_x = I \tag{B.1}$$

The problem is to choose n_x so that the rate at which undetected T pass through the inspection point is minimized (although note, in practice, that the values of p_x, N_x and I are such that this approach would be aiming to minimize something that is very small). Assuming perfect detection (for simplicity), the expected number of T's from x that are undetected in unit time is

$$E(n_x) = (N_x - n_x)p_x \tag{B.2}$$

and the probability that no T from x passes through the inspection point undetected is:

$$p(n_x) = (1 - p_x)^{N_x - n_x} \approx \exp[-E(n_x)] \tag{B.3}$$

With limited resources, one would like to minimize $E(n_x)$ or equivalently maximize $p(n_x)$, for each source. This is, of course, achievable when everything is inspected (i.e., $n_x = N_x$ where $E(n_x) = 0$ and $p(n_x) = 1$, regardless of p_x). In practice, there are constraints on the total number of items that could be inspected. Hence we would need to minimise $\prod p(n_x)$ or minimise $\sum E(n_x)$ subject to the resource constraint.

We will consider the case of two sources, $x = A$ or B, with constraints

$$n_A + n_B = I; \quad 0 \le n_x \le I \tag{B.4}$$

We will assume that $I < N_x; x = A, B$ in the following development.

B.2 Stratified/variable probability sampling

Initially, with little or no knowledge about p_x, the best that one can do is to choose

$$n_A = \frac{N_A I}{N_A + N_B} \quad \text{and} \quad n_B = \frac{N_B I}{N_A + N_B} \tag{B.5}$$

B.3 Knowing something about p_x

Consider the objective of management to minimize the number of undetected failures, which can be achieved by maximizing the number of inspections in areas where the risks are greatest, at the same time acquiring information and maintaining deterrence on less critical pathways. This implies that the most effective strategy will be to allocate effort somewhat in proportion to the magnitude of a threat, so that minor threats receive small (but non-zero) attention, and more important threats receive greater attention, without suborning all resources.

The magnitude of a threat may be expressed as the product of the number of objects arriving from a source, multiplied by the probability that each object contains a threat. That is, the magnitude of a threat associated with a pathway of entry is represented by the number of threats expected to arrive by that pathway. Using the terms defined previously, namely

p_x is the probability an object from source x contains a threat, T,

N_x is the number of objects from x that pass through an inspection point in some unit time interval,

n_x is the number of objects from x inspected in the unit time interval for the common threat, T.

we measure the allocation of effort to a pathway, relative to the threat associated with that pathway, by

$$R_x = \frac{n_x}{N_x p_x} \tag{B.6}$$

The solution that gives the same relative effort to each group, given two pathways of entry (two sources, A and B), is

$$\min_{(n_a, n_b)} (\max(R_A, R_B)) \tag{B.7}$$

(Recall that $n_a + n_b = I$, which is assumed to be fixed). The functions for R_A and R_B intersect when

$$\frac{n_A}{N_A p_A} = \frac{n_B}{N_B p_B} \tag{B.8}$$

giving optimal solutions

$$n_A = I \frac{p_A N_A}{p_A N_A + p_B N_B} \quad \text{and} \quad n_B = I \frac{p_B N_B}{p_A N_A + p_B N_B} \tag{B.9}$$

This solution can be seen graphically in an example in figure B.1.

B.4 Minimizing the total number of undetected T's

If one has estimates for p_x then one should seek to minimize

$$E(n_A, n_B) = E(n_A) + E(n_B) = (N_A - n_A)p_A + (N_B - n_B)p_B \tag{B.10}$$

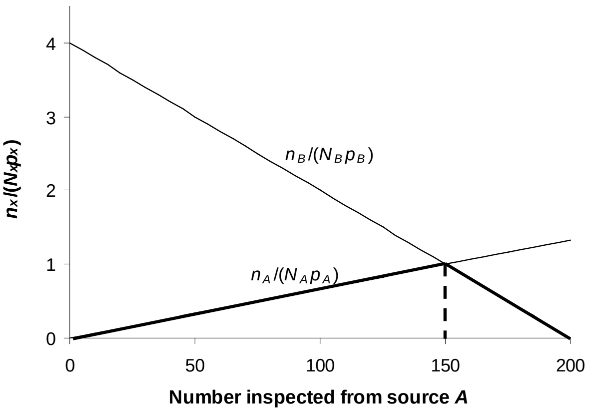

Figure B.1: Example solution that allocates inspection resources based on the expected number of threats. $n_x/(N_x p_x)$ represents the allocation of resources in proportion to the magnitude of a threat (or in proportion to the number of threats expected to arrive by a specified pathway). The two lines represent the curves for samples allocated to sources A and B. In this example, equal volumes of trade arrive from the two sources, but the probability of a threat from source A is 0.15 and from source B is 0.05. The most effective allocation occurs according to formula B.9, in this case, when 150 inspections out of 200 are allocated to source A and the balance, 50 samples, are allocated to source B, indicated by the heavy dashed vertical line.

This is achieved (from B.5) when

$$n_A = \left\{ \begin{array}{ll} I, & p_A > p_B \\ I/2, & p_A = p_B \\ 0, & p_A < p_B \end{array} \right\} \tag{B.11}$$

We deem this unsatisfactory because no further knowledge would be gained through inspections about the smaller of p_A and p_B, and there would be no deterrence for the source that was thought to be less risky.

B.5 Balancing risks

Recall our earlier example of a commodity entering a port from two sources (countries or suppliers), labelled A and B. Different volumes of trade are expected over some fixed period (a week, a month or a year). Assume a fixed budget that allows a total of I inspections, allocated among the sources, n_A and n_B respectively (that is, $n_A + n_B = I$). The risk-sensitive allocation of resources (the number of inspections) among the sources may be calculated by;

$$n_A = I\frac{F_A N_A}{F_A N_A + F_B N_B} \quad \text{and} \quad n_B = I\frac{F_B N_B}{F_A N_A + F_B N_B} \tag{B.12}$$

where

- N_A and N_B are the *predicted* volumes of trade (the number of units) expected from sources A and B respectively, and
- F_A and F_B are the *predicted failure risks* that an individual unit from A or B will contain at least one threat.

Figure B.2 shows a situation in which the failure risk from source A is 0.15 and in source B is 0.05. There are a total of 200 inspections to be allocated among the two sources. When equal volumes of trade arise from each source (*i.e.* the proportion of trade from A is 0.5), as indicated by the dashed vertical line, then the best strategy is to allocate 150 inspections to the higher risk source (A) and 50 inspections to the lower risk source (B).

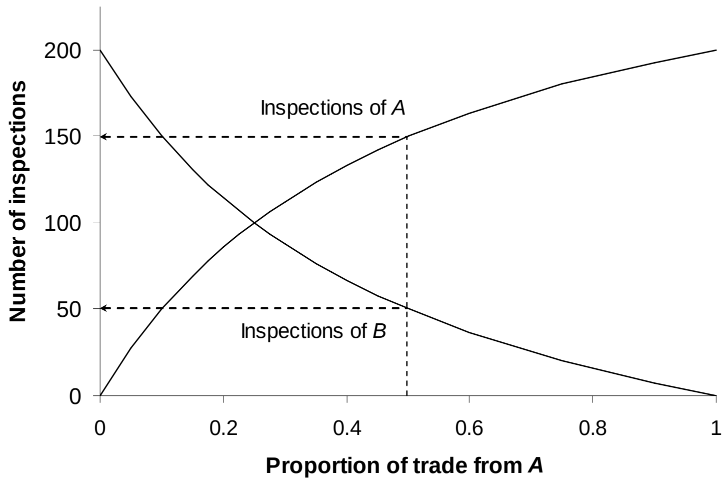

Figure B.2: Allocation of inspection resources among two sources. In this example, the failure risk from A is 0.15 and from source B is 0.05. The decision involves allocating a fixed number of inspections (200) between the two sources. When volumes of trade are equal (indicated by the vertical dashed line), the most effective solution is to inspect 150 lots from source A and 50 lots from source B, indicated by the two horizontal lines.

This approach fails because it assumes a fixed total of I inspections, and it leads to unsatisfactory results in the following scenario. Given $N_A = 200$, $F_A = 1$, $N_B = 2000$, $F_B = 0.1$, and $I = 200$, we would allocate the resources equally, but it makes much more sense to exhaust A before starting on B (notwithstanding deterrence and monitoring etc.)

Appendix C

Other Analysis Details

This appendix provides intermediate analysis details and tables of possible interest.

C.1 Aircan surveillance

Table C.1: Number of failures of aircans, presented by region and nationally.

	2003	2004	2005	2006	2007	Total
Brisbane	2	5	6	7	58	78
Far North	9	33	44	21	33	140
NSW	79	379	283	98	137	976
NT	0	0	1	0	0	1
SA	44	33	112	30	59	278
VIC	74	54	56	22	24	230
WA	0	1	7	0	0	8
National	208	505	509	178	311	1711

Table C.2: Number of external inspections of aircans, presented by region and nationally.

	2003	2004	2005	2006	2007	Total
Brisbane	13835	33137	37626	39789	37743	162130
Far North	5834	10922	3694	3152	2957	26559
NSW	79861	191295	196130	203259	207764	878309
NT	450	614	458	415	0	1937
SA	5847	13234	13789	15620	17510	66000
VIC	61303	121982	92971	92980	91491	460727
WA	24675	34436	13946	13344	14067	100468
National	191805	405620	358614	368559	371532	1696130

Table C.3: Predicted 95% failure risk for external inspection of aircans, presented by region and nationally, expressed as a percentage. TAS had no inspections in the period.

	2003	2004	2005	2006	2007
Brisbane	0.0400	0.0297	0.0297	0.0314	0.1897
Far North	0.2582	0.3986	1.5137	0.9394	1.4702
NSW	0.1186	0.2154	0.1589	0.0568	0.0757
NT	0.4257	0.3122	0.8500	0.4615	0.0000
SA	0.9569	0.3290	0.9457	0.2567	0.4152
VIC	0.1456	0.0551	0.0746	0.0332	0.0363
WA	0.0078	0.0113	0.0896	0.0144	0.0137
National	0.1214	0.1339	0.1526	0.0545	0.0918

Table C.4: Predicted cumulative 95% failure risk for external inspection of aircans, presented by region and nationally, expressed as a percentage. TAS had no inspections in the period.

	2003	2004	2005	2006	2007
Brisbane	0.0400	0.0266	0.0237	0.0229	0.0577
Far North	0.2582	0.3207	0.5003	0.5298	0.6043
NSW	0.1186	0.1823	0.1684	0.1324	0.1171
NT	0.4257	0.1803	0.2564	0.2015	0.2015
SA	0.9569	0.4847	0.6468	0.5039	0.4643
VIC	0.1456	0.0806	0.0751	0.0625	0.0556
WA	0.0078	0.0066	0.0189	0.0160	0.0137
National	0.1214	0.1269	0.1339	0.1104	0.1049

C.2 White rice surveillance

Table C.5: Number of failures of white rice, presented by region and nationally.

	2003	2004	2005	2006	2007	Total
NSW	0	1	0	0	0	1
SA	0	0	0	1	0	1
VIC	0	1	0	0	0	1
WA	0	0	0	0	0	0
Other	0	0	0	0	0	0
National	0	2	0	1	0	3

Table C.6: Number of external inspections of white rice, presented by region and nationally.

	2003	2004	2005	2006	2007	Total
NSW	15	33	58	41	35	182
SA	5	16	19	16	12	68
VIC	12	36	29	29	39	145
WA	11	27	31	38	39	146
Other	7	3	9	3	6	28
National	50	115	146	127	131	569

Table C.7: Estimated 95% failure risk for inspection of white rice, presented by region and nationally, expressed as a percentage. *Other* represents NT, QLD, SEQLD, and TAS combined.

	2003	2004	2005	2006	2007
NSW	11.8	11.2	3.2	4.5	5.3
SA	30.6	11.1	9.5	22.0	14.5
VIC	14.5	10.4	6.4	6.4	4.8
WA	15.7	6.8	6.0	4.9	4.8
Other	23.2	44.4	18.7	44.4	26.4
National	3.8	4.7	1.3	3.0	1.5

Table C.8: Estimated cumulative 95% failure risk for inspection of white rice, presented by region and nationally, expressed as a percentage. *Other* represents NT, QLD, SEQLD, and TAS combined.

	2003	2004	2005	2006	2007
NSW	11.8	7.9	3.6	2.6	2.1
SA	30.6	8.6	4.7	6.8	5.6
VIC	14.5	7.9	5.0	3.6	2.7
WA	15.7	4.9	2.7	1.8	1.3
Other	23.2	17.1	9.5	8.3	6.6
National	3.8	3.3	1.8	1.6	1.2

C.3 Aircargo (HVLV) surveillance

Table C.9: Number of failures of HVLV, presented by region and nationally.

	2003	2004	2005	2006	2007	Total
Brisbane	32	55	27	86	140	340
Far North	2	5	2	2	4	15
NSW	418	524	365	459	304	2070
NT	0	0	0	2	2	4
SA	0	0	0	0	0	0
VIC	24	44	235	99	16	418
WA	7	19	38	38	17	119
National	483	647	667	686	483	2966

Table C.10: Number of external inspections of HVLV, presented by region and nationally.

	2003	2004	2005	2006	2007	Total
Brisbane	65332	126309	125586	128538	143778	589543
Far North	3530	7485	7068	6975	3844	28902
NSW	774075	1652697	1682627	1635539	1754577	7499515
NT	0	0	13	3338	5906	9257
SA	7696	20299	25196	35447	38917	127555
VIC	146448	300796	302516	439452	557516	1746728
WA	50162	112381	123676	146521	151866	584606
National	1047243	2219967	2266682	2395810	2656404	10586106

Table C.11: Predicted 95% failure risk for external inspection of HVLV, presented by region and nationally, expressed as a percentage. TAS had no inspections in the period.

	2003	2004	2005	2006	2007
Brisbane	0.0649	0.0541	0.0292	0.0796	0.1117
Far North	0.1567	0.1314	0.0783	0.0793	0.2199
NSW	0.0585	0.0340	0.0236	0.0303	0.0190
NT	0.0000	0.0000	13.4893	0.1657	0.0937
SA	0.0250	0.0095	0.0076	0.0054	0.0049
VIC	0.0226	0.0186	0.0864	0.0265	0.0043
WA	0.0249	0.0243	0.0398	0.0336	0.0164
National	0.0497	0.0311	0.0313	0.0305	0.0196

Table C.12: Predicted cumulative 95% failure risk for external inspection of HVLV, presented by region and nationally, expressed as a percentage. TAS had no inspections in the period.

	2003	2004	2005	2006	2007
Brisbane	0.06491	0.05397	0.04181	0.05033	0.06300
Far North	0.15672	0.11343	0.08333	0.07017	0.07781
NSW	0.05848	0.04094	0.03328	0.03196	0.02861
NT	99.38442	99.38442	13.48929	0.16508	0.09136
SA	0.02495	0.00686	0.00361	0.00217	0.00151
VIC	0.02265	0.01848	0.04437	0.03667	0.02592
WA	0.02491	0.02184	0.02734	0.02766	0.02361
National	0.04968	0.03631	0.03375	0.03236	0.02887

C.4 Imported Food

Table C.13: Summary of imported food data. Those columns with labels subscripted with p refer to the profile-construction data, those with v refer to the independent validation data. n is the number of inspections, k is the number of failures, $\hat{p}$ is the estimated failure rate, and $\hat{p}$ is the estimated failure rate.

Test	n_p	k_p	$\hat{p}_p$ (%)	n_v	k_v	$\hat{p}_v$ (%)
1 3 Dichloropropanol	86	0	0.0	65	0	0.0
3Monochloropropan1,2Dl	87	1	1.1	65	0	0.0
Aflatoxin	455	16	3.5	355	14	3.9
Artificial Sweeteners	19	1	5.3	0	0	0.0
B Cereus	39	7	17.9	267	4	1.5
Bse Cert Check	600	5	0.8	703	2	0.3
Cadmium	411	21	5.1	352	16	4.5
Carbadox	39	0	0.0	0	0	0.0
CheckLablFrAddtv(nTst)	10	3	30.0	6	0	0.0
Chloramphenicol	134	0	0.0	105	0	0.0
Coliforms	2	0	0.0	0	0	0.0
Colour Screen	105	3	2.9	87	1	1.1
Commercial Sterility	237	8	3.4	0	0	0.0
Composition	173	73	42.2	138	47	34.1
Domoic Acid	65	0	0.0	54	0	0.0
E Coli	947	49	5.2	737	18	2.4
Ethylene Chlorohydrin	291	2	0.7	327	1	0.3
Histamines	909	16	1.8	852	19	2.2
Inorganic Arsenic	1	1	100.0	1	1	100.0
Labelling	11901	615	5.2	9979	574	5.8
Lead	70	5	7.1	52	0	0.0
Listeria Monocytogenes	908	12	1.3	663	10	1.5
Malachite Green	33	0	0.0	36	0	0.0
Mercury	559	16	2.9	285	7	2.5
Nitrofurans	172	12	7.0	142	8	5.6
Officer To Assign Test	6	0	0.0	0	0	0.0
OystersExJapan/Korea	1	1	100.0	1	1	100.0
Patulin	8	0	0.0	0	0	0.0
Pesticides	2190	13	0.6	2356	8	0.3
Ph	441	0	0.0	1	0	0.0
Psp Toxin	65	0	0.0	54	0	0.0
Radiation	1	1	100.0	0	0	0.0
Salmonella	2185	24	1.1	1599	12	0.8
Standard Plate Count	295	8	2.7	216	10	4.6
Staph Enterotoxin	840	1	0.1	0	0	0.0
Staphylococci	248	2	0.8	220	0	0.0
Streptomycin	2	0	0.0	3	0	0.0
Sulphonamides	2	0	0.0	3	0	0.0
Sulphur Dioxide	146	4	2.7	100	1	1.0
Tetracyclines	3	0	0.0	5	0	0.0
Vibrio Cholerae	187	2	1.1	133	0	0.0
Visual	11322	13	0.1	9503	16	0.2
OVERALL	36195	935	2.6	29465	770	2.6

Appendix D

Remaining Questions, Issues and Opportunities

1. How do we adjust for time of year (*e.g.* European / Asian gypsy moth)?
2. How do we deal with language uncertainty, *e.g.*, "packaging", "wool tops", ambiguous terms in ICON?
3. How do we ensure the consistency and quality of *Incidents* data?
4. Profiles are a set of rules that may lead to an inspection. How do we measure and improve the performance of profiling: leakage surveys, snapshot surveys, regression techniques (boosted regression trees?), ROC curves? Can profiling reduce interventions, through more efficient resource allocation? An Internal Audit Review of AQIS profiling will be complete in January/February 2008. Is it available?
5. Data Review Project 9 Sarah Gossling, Canberra. Is it available?[1]
6. Is it possible for a data management system to provide corporate memory for trends and differences in trends over time, between regions, by QAP, importer, commodity and kind of event?
7. How do we incorporate agreed test reliabilities (IPPC 95% chance of finding a pest)? *e.g.* Fresh Produce Manual specifies a sample size of 600, including 450 under 10x magnification and 60 under 100x magnification. This may mean sampling 600 fruit in 4 containers.
8. How should we ensure consistency of effort (relative to risk), especially for general produce other than commodities. Should we reconcile these efforts with the Australian Standard for Sampling?

Opportunities exist for making decisions easier for the individual sections within the regional office, should sufficient flexibility be available.

1. Connecting Incidents, AIMS, SAC, and ICE databases would allow reports of infraction detection rates and trends by commodity, importer, broker, AFAS licence number, QAP, etc. Those reported trends could be used to inform more precise profile construction, guide resource allocation for ICE, BAS, or QAP audits, identify opportunities in AQIS communications, and focus attention on persistent problem areas for potentially stronger or better-informed ameliorative actions.

[1] WA *claims* to have forwarded.

2. Tools such as confusion matrices, logistic regression, and simple smoothers, could assist in identifying patterns from the available data.

3. A statistical model, for example belonging to the objective Bayes family, could estimate the probability of detecting an infraction, and informally balance this probability across commodities, importers, and brokers. Precedents exist for this kind of piecemeal (one commodity, one broker, or one importer at a time) modelling.

4. Fitting models that cover many commodities, importers, and brokers, and borrow strength from each different area would support more reliable conclusions about the available space of actions. The goal here would be to predict the probability of detecting an infraction as a function of various inputs, and using the modelled probability to guide resource investment, conditional on collecting sufficient data to be able to continue updating the model robustly. The system may support strategies in the National Resource Model which tracks staff resources across the country on a periodic basis.

Risk assessments should consider country, broad geographic region and commodity, and provide adaptable inspection systems so that surveillance can change as hazards change over time, and can target high risk pests and diseases. The sampling rate needs scientific justification and verification.

AQIS collects a large amount of useful data. Some basic data are common to all commodities: inspection type, result (fail/pass), commodity/non-commodity, concern (list of types), supplier, commodity (list of types), source (country), entry mode.

1. The process for adding an attribute to the data base is not always clear, and feedback to regions on advice that the regions give to Canberra may need to be clarified.

2. Training can improve the efficiency of fruit inspection substantially. For example, discussions with staff suggest detection rates of insects in fruit from the United States increased 32% to 67% to 79.8% with two training periods (I. Hanson, pers. comm.).

3. In many circumstances, inspectors have a reliable prior expectation of detecting pests or diseases: on grapefruit and lemons from the USA, detection probability is close to 0. For oranges from the USA, detection probability is close to 100%.

4. Passive surveillance during inspections might feed back to ICE and IPU more formally or be used somehow as auxiliary information. Passive surveillance is a source of regular, but inconsistent information.

5. If pests and diseases cluster within containers then sampling fewer commodities from more containers will increase the probability of interception.

6. The ideal relationship between AQIS and its partners is cooperative, not adversarial. Therefore the ideal system will be one that monitors for and discourages infractions, and supports and rewards appropriate respect for the nation's quarantine needs and values. However, AQIS operations sometimes force them to effectively punish companies or citizens for good acts – *e.g.* the clients report quarantine problems and then are required to pay for inspection.

7. If air cargo is x-rayed by ACS, it is not recorded on the good's import record. Therefore decisions are made about the next step without knowing that the good has been x-rayed.

8. Reports of trends of QAP failures by premises class, and/or failure type, and compliance agreement audit failures would permit pro-active management. This may be implemented with spreadsheet models to assist officers on ground to update judgements and allocate effort.

9. Reports of trends of QAP failures by premises class, and/or failure type, and compliance agreement audit failures, movements, charges, client details would support more consistent decision-making.
10. Reports of risk trends by commodity, source area, importer, for each region/port, would improve consistency of practice across regions and detect changes in processes associated with suppliers, source countries, commodities. For example, Asian gypsy moths are in their breeding cycle in September. The pest should be anticipated in shipping containers from Russia over the period September – October.
11. Reports generated using dash boards, 6-Sigma style reporting systems, compatible with EXCEL are preferred because the information may be locally accessed and manipulated for local reporting requirements.

Flexible inspection systems (such as, for example, a rating scale) could apply to QAPs. Reports on trends would be useful against type of premises, region (State), and other criteria.

Appendix E

Change Log

22 December 2008 Suggestions from Moshe Sniedovich.

- Fix typos in Appendix B.
- Absorbed material in previous Section B.4 Information Gap/max-min/worst-case into Section B.3.

12 December 2008 Suggestions from Greg Hood.

- Added symbols to graphs to facilitate interpretation in monochrome printings.
- Clarified confusion between "sources" and "commodities" in Section 4.5.1, by renaming all the former into the latter.