Report Cover Page

ACERA Project
1101C
Title
Plant Quarantine Inspection and Auditing across the Biosecurity Continuum
Author(s) / Address(es)
Andrew Robinson, Australian Centre of Excellence for Risk Analysis Brendan Woolcott, DAFF Peta Holmes, DAFF Adam Dawes, DAFF Jessica Sibley, DAFF Leisa Porter, DAFF James Kirkham, DAFF
Material Type and Status
Final Report
Summary

- **Background:** Previous collaborative ACERA/DAFF work has identified a suite of imported plant-product pathways that could be managed with a risk-based approach using statistical tools such as the Continuous Sampling Plan, CSP.
- **Overview:** This project extends the earlier work by developing statistical tools that will enable automated identification of risky and safe pathways, and pathways with characteristics or patterns that might provide a good match for CSP. We studied fresh and dried dates, medium-risk nursery stock (MRNS), and medium-risk plant-based stockfeed (PBS).
- **Outcomes:**
  - The inspection of fresh as well as dried dates can be managed using the CSP–3 tool under the single tariff for dates, splitting the pathway by *importer*.
  - There is considerable variation between the regions on the rate at which contamination is detected within the MRNS pathway.
  - Altering the cutoff number of contaminated plants that constitutes a quarantine failure in MRNS has a minimal effect upon the pathway-level contamination rate, and further consideration is not justified.
  - Analyzing the patterns of contamination in PBS is very challenging because it has so far proved impossible to develop an inspection database owing to the lack of a definitive identifier for the pathway.
  - Several pathways in the fruit tariff chapter are identified as low or high risk by use of simple data mining tools.

- **Recommendations:**
  - Rather than attempting to develop and support the analytical skills required to maintain and update CSP–3 inspection algorithms and data-mining activities, the pathway manager should seek support from a provider such as ABARES.
  - Use CSP for risk-based management of all consignments entering under the dates tariff, whether fresh or dried, splitting the pathway by importer.
  - For other pathways being assessed for CSP, DAFF should consider splitting the pathways by importer.
  - Continue data-mining activities on fruit inspection data (presently underway as ACERA Project 1206F).
  - Improve border data capture of permit data for PBS.

ACERA Use Only	Received By:	Date:
	ACERA / AMSI SAC Approval:	Date:
	DAFF Endorsement: ( ) Yes ( ) No	Date:

Plant Quarantine Inspection and Auditing across the Biosecurity Continuum.

ACERA 1101C Final Report

Andrew Robinson, Australian Centre of Excellence for Risk Analysis
Brendan Woolcott, DAFF
Peta Holmes, DAFF
Adam Dawes, DAFF
Jessica Sibley, DAFF
Leisa Porter, DAFF
James Kirkham, DAFF

September 3, 2013

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) and the University of Melbourne.

The authors thank Christina Ashton for her valuable contribution. The authors are also grateful to Tony Arthur, Simon Barry, Matthew Smyth, and an anonymous reviewer for their detailed and constructive reviews, and to Enrico Perotti and Louise Van Meurs for their enthusiastic support and very useful suggestions.

Contents

List of Tables

List of Figures

Table 2: Table of definitions used throughout the text.

Term	Definition
Assumption	An *assumption* is a statement of belief about some condition that can't be proven. Assumptions can be considered reasonable or unreasonable in the face of evidence. Assumptions are made in order to use or connect statistical arguments.
CSP	*CSP* stands for continuous sampling plan. A CSP is a method for determining whether or not to inspect a consignment, based on the recent inspection history of the pathway, and some parameters that the pathway manager sets. More information about the CSP approaches is available in Section 2.1.
Enhanced Inspection	For the purposes of this report, *enhanced inspection* is to inspect all of the consignments on a pathway, where the inspection is carried out according to the usual work instructions. Thus, enhanced inspection of the fresh oranges pathway would involve the usual 600–unit sample applied to all of the consignments of fresh oranges. Also see *monitoring*, below.
Leakage	*Leakage* is the amount of undetected biosecurity risk material that passes through an intervention point.
Monitoring	We will use *monitoring* to describe the inspection of less than 100% of the consignments of a pathway. Thus, monitoring the fresh oranges pathway would involve the usual 600–unit sample applied to only some of the consignments of fresh oranges; the remainder would be released on documentation. Monitoring would be performed for pathways that are considered low risk by the pathway manager. Also see *enhanced inspection*, above.
Pathway failure (PF)	A *pathway failure* will be any kind of non-compliance associated with a consignment on a pathway, including failures that do not necessarily represent a biosecurity risk. For example, inadequate documentation for a consignment is a pathway failure, as is contamination by a pest or disease.
Quarantine failure (QF)	A *quarantine failure* will be non-compliance that is a biosecurity risk associated with a consignment on a pathway. For example, contamination by a pest or disease is a quarantine failure, but inadequate paperwork is not.
Sampling	We will use *sampling* to refer to the selection of individual units within a consignment as part of the usual inspection procedure. For example, the selection of 600 oranges from a consignment will be referred to as sampling.

1

Executive Summary

This report summarizes ACERA project 1101C, *Quarantine Inspection and Auditing Across the Biosecurity Continuum*. It is written to satisfy the first and third deliverables of the 1101C project, namely coverage of progress on the following sub-projects.

- Case Studies (Chapter 3) — we provide the following deliverables.
  - A re-analysis of dates inspection data (now including fresh as well as dried dates).
    - We conclude that a risk-based approach to a reduced inspection regime is reasonable, and that the CSP–3 algorithm is suitable. We note that the new analysis is easier to operationalize than that reported for dried dates in Robinson et al. (2012) because it relies solely on the dates tariff (8041000), and is more reliable because it is based on a larger number of inspections. The pathway should be stratified by importer. The risk-based approach to inspection for dates (dried and fresh) was implemented in July 2012 (see Section 3.1).
  - An analysis of medium-risk nursery stock (MRNS) inspection data.
    - This analysis concluded that (i) there was a substantial and statistically significant difference between the interception rates of the different regions, and that (ii) when contamination is detected on plant consignments, it is very often detected on a large proportion of the plants, so there is no operational value in considering an increase the tolerance from zero contaminated plants to one or two contaminated plants per consignment (see Section 3.2).
  - An analysis of plant-based stockfeed inspection data.
    - This analysis was unable to be done because it was deemed too difficult to develop a dataset that contained the inspection results for plant-based stockfeed only. The difficulty lay in the fact that DAFF's databases are *transactional*, and focus on recording information to mitigate and manage the potential biosecurity risk of the consignments crossing the border. Hence the databases are not well suited to retrospective analyses of this nature (see Section 3.3).
- Refinement of the statistical tools — progress has been made, focusing on inspection data from consignments of fruit, which correspond to Chapter 8 of the tariffs.

  - We classified the fruit pathway by tariff into three risk groups, based on arbitrary but reasonable cutoffs of their interception rate and their current sampling rate. We found that nuts and dried fruits were generally low risk, and fresh fruits were higher risk (see Chapter 4).

- Development of data-mining tool (Chapter 4) — progress has been made, focusing on inspection data from consignments of fruit, which is Chapter 8 of the tariffs. This deliverable has now been transferred to ACERA project 1206F.
- Determination of training needs (Chapter 5) — DAFF and ACERA are working with ABARES to enable ABARES to assume maintenance of the tools and algorithms. Therefore the training needs will be covered by ABARES internal processes.
- Identification of processes needed to measure biological risk (Chapter 5) — this deliverable *could not be achieved* as it requires the outcomes of ACERA Project 1101E, *Sampling for Invasives.*
- Development of a risk–return analysis for the imported plant pathways using available data and provision of recommendations for auditing, quarantine inspection intensities, data reassessment timelines and training of Plant Import Operations Staff — this deliverable requires deeper analysis than was possible during this project and forms part of one of the proposed CEBRA projects.

2

Introduction

This report concludes ACERA project 1101C, *Quarantine Inspection and Auditing Across the Biosecurity Continuum*. It is written to satisfy the first and third deliverables of the 1101C project, namely coverage of:

- Case Studies (Chapter 3)
- Refinement of the statistical tools (Chapter 4)
- Development of data-mining tool (Chapter 4)
- Determination of training needs (Chapter 5)
- Identification of processes needed to measure biological risk (Chapter 5)

We begin with a review of the statistical algorithm that was presented in Robinson et al. (2012).

2.1 Continuous Sampling Plans

The effective review of an inspection regime requires employment of a monitoring technique. Here we introduce the continuous sampling plan (CSP) family of algorithms as the recommended pathway monitoring technique for imported-plant product pathways. The original CSP algorithm of Dodge (1943) is now called CSP–1. CSP–3, which we advocate here, was introduced by Dodge and Torrey (1951), along with CSP–2.

The basic premise of each of the CSP inspection designs that we reviewed is that a pathway is either being *monitored* or undergoing *enhanced inspection* at any given time, and the decision that the inspectorate must make is: how to use the inspection history of the pathway. Typically, the inspection regime will adopt the following very general pattern:

1. start in enhanced inspection mode,
    - inspect all consignments until the inspection history reaches a given condition, say C,
2. switch to monitoring mode:
    - inspect at a specific monitoring rate f until the inspection history reaches a different given condition, say, M, then switch back to enhanced inspection mode.

We now briefly review two of the simpler inspection algorithms, CSP–1 and CSP–2, before providing a review of the recommended approach, CSP–3.

2.1.1 CSP–1

CSP–1 is the simplest of the three algorithms. Referring to the pattern immediately above,

- C is satisfied by the observation of i successive compliant consignments.
- M is satisfied by any observation of non-compliance while the pathway is in the monitoring mode.

2.1.2 CSP–2

CSP–2 is the same as CSP–1, except

- M is satisfied by two observations of non-compliance within k successive consignments while the pathway is in the monitoring mode. It is usual to take $k = i$.

2.1.3 CSP–3

For CSP–3, the inspection algorithm is as follows,

- C is satisfied by the observation of i successive compliant consignments.
- M is satisfied by two observations of non-compliance that are within k inspections of one another. Usually, $k = i$.

Note that *in addition to the above prescription*, when any non-compliance is detected, the next four consignments will all be inspected, i.e., the monitoring process will be replaced by inspecting each of the next four consignments. The choice of four is arbitrary, but standard for CSP–3. This shift is temporary and is done regardless of whether the non-compliance is the first or subsequent failure detected in any number of consignments.

To summarize CSP–3, if the system is in monitoring mode, then the consignments are being randomly selected for inspection at rate f. If a non-compliant consignment is intercepted, then the next four consignments are inspected, and if they are all compliant, the system returns to monitoring. If another non-compliant consignment is found within i inspections of the previous non-compliance, then the system switches to enhanced inspection mode.

3

Case Studies

This chapter details three case studies that build on the collection reported in Robinson et al. (2012), namely, fresh and dried dates, and medium-risk nursery stock (MRNS). Each case study applies the same basic methodology as detailed in Robinson et al. (2012), with some modification, as follows. The first study includes fresh as well as dried dates in the consignments used for the analysis. The second focuses on analysis of the patterns of inspection outcomes to determine (i) whether there seems to be any statistical evidence to suggest that the contamination rates were different in the different regions, and (ii) to determine whether there might be any efficiency advantage to changing the definition of a failure.

3.1 Dates, Dried & Fresh

3.1.1 Introduction

We re-analyse the historical inspection data for dates, using algorithms developed in Robinson et al. (2012). This analysis differs from that reported previously in that it includes fresh as well as dried dates. The methodology is otherwise identical. Therefore the relevant contextual information can be found in that report.

The motivation for this case study was that pathways can only be unambiguously defined using data that are reliably available at the time of processing. There is only one tariff for dates, whether they are fresh or dried, namely 08041000. In the previous analysis we used the *goods description* field to try to distinguish between dried and fresh dates, but this is not entirely satisfactory because the goods description does not necessarily reliably capture the treatment of the goods. If it could be demonstrated that the whole pathway defined by the dates tariff, whether dried or fresh, was suitable for a reduced adaptive inspection regime then there would be no need to try to distinguish between the dried and the fresh dates.

It is important to note that we are also assuming that the types of biological risks that are associated with the dried and fresh dates are essentially equivalent.

In this analysis we have assumed that inspection effectiveness is 90%, that is, that if a consignment is contaminated then the probability of detecting the contamination is 90%. This assumption is based on discussions with the pathway manager. Generally this figure should be estimated using a *leakage survey*, however, leakage surveys are fully implemented in only a small number of DAFF pathways, and work is underway at the time of writing to implement these techniques across a wider range of DAFF activities. At present, making such a strong assumption is unavoidable.

Analytic Timespan

It is inevitable that there is a trade-off between data timeliness and the observation count. The reporting period was constrained to inspections from July 2008 to January 2012 (totaling 2349 consignments), in order that the results reflect the current status of the pathway as much as possible.

Pathway Summary

The pathway is summarized in Figure 3.1.

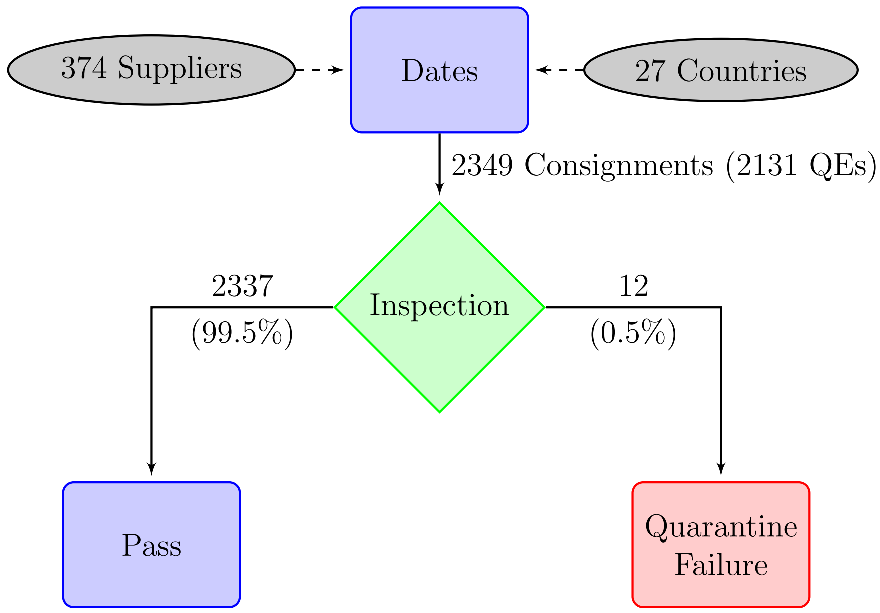

Figure 3.1: Dates consignments flow chart with statistics for July 2008–January 2012. A quarantine failure was recorded for consignments with a detection of quarantine concern, such as insect, pathogen, or contamination.

3.1.2 Analysis

Summaries

This section provides a statistical overview of the data. The full dataset comprises 3983 consignments with record creation dates ranging from December 2005 to January 2012, and comprises entries from 38 countries and 573 suppliers.

A smoothed plot of the quarantine failure rate against time is presented in Figure 3.2. The figure shows a failure rate descending smoothly from just under 1% to less than 0.5%. The failure rate for the entire period was 0.55%, and for the analysis period (everything after June 2008) was 0.51%.

The pattern of quarantine failure counts by country and supplier is presented in Table 3.2. The statistics in Table 3.3 summarize the inspection data for those countries with at least five consignments during the key time period. Table 3.4 reports the number of

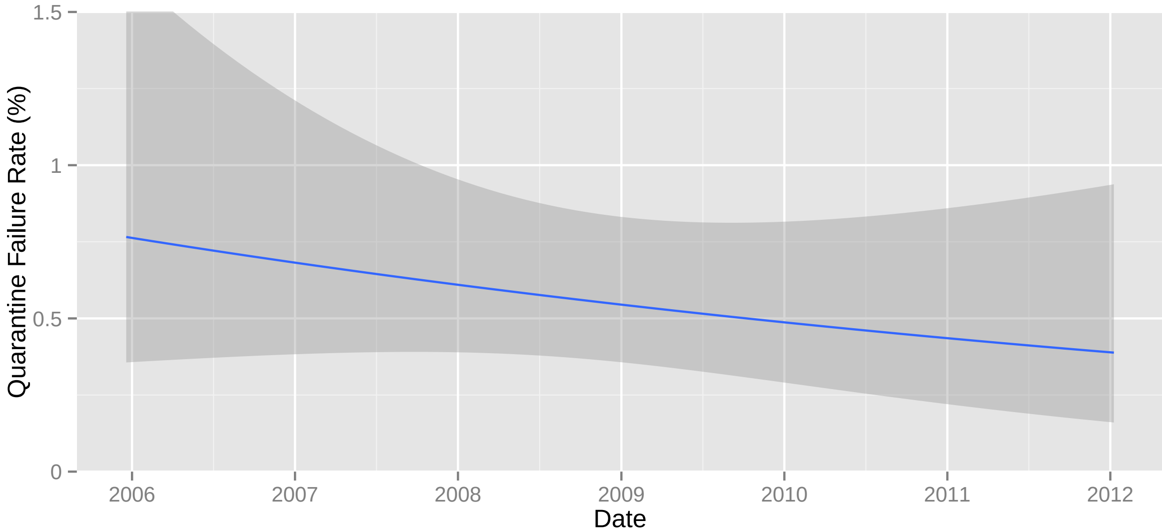

Figure 3.2: Quarantine failure rates (%) smoothed by date, with a 95% confidence interval (shaded region) added. The width of the shaded region indicates the uncertainty of the line, which becomes narrower as the (nearby) sample size increases.

Table 3.1: Pattern of inspections and quarantine failure counts by year for dried and fresh dates. *Count* is the number of consignments imported during the study period, *QF* is the count of consignments with contamination of quarantine interest, *QF %* is the percentage of consignments with contamination of quarantine interest, and *Tonnage* is the total tons of product imported during the study period. Note that 2005 and 2012 are not full years.

Year	Count	QF	QF %	Tonnage
2005	6	0	0.0	44
2006	695	5	0.7	6,261
2007	653	5	0.8	6,642
2008	554	4	0.7	4,998
2009	592	3	0.5	4,847
2010	732	0	0.0	6,459
2011	743	5	0.7	6,876
2012	8	0	0.0	70

consignments with and without the word `FRESH` in the goods description. Table 3.5 summarizes the inspection data for the suppliers with at least 15 consignments, and Table 3.6 summarizes the inspection data for the suppliers with at least 15 consignments. Note the small number of quarantine fails that appear in each of the last two tables: most of the fails are occurring among the consignments from the low-volume suppliers and importers.

Table 3.2: Pattern of recent quarantine failure counts by country, by supplier, and by importer, for dried and fresh dates. The data cover all inspections between July 1 2008 and 9 January 2012. Note that the columns are separate and do not cross-refer.

Failures	Countries	Suppliers	Importers
0	20	584	396
1	4	8	4
2	2	2	4
3	0	0	0
4	1	0	0

Table 3.3: Summary statistics by country for dried and fresh dates. *Count* is the number of consignments imported during the study period, *QF* is the count of consignments with contamination of quarantine interest, *QF %* is the percentage of consignments with contamination of quarantine interest, and *Tonnage* is the total tons of product imported during the study period. The *Suppliers* and *Importers* columns report the number of suppliers and importers, respectively, that have exported (imported) from each country during the time period. The data cover all inspections between July 1 2008 and 9 January 2012. We only include those countries with five or more consignments during the time period.

Country	Count	QF	QF %	Tonnage	Importers	Suppliers
Iran Islamic Republic Of	693	0	0.0	12,596	46	56
China	552	0	0.0	743	62	97
United Arab Emirates	269	2	0.7	1,001	44	50
United States	240	4	1.7	1,983	24	29
Turkey	100	1	1.0	1,353	7	4
Saudi Arabia	85	1	1.2	127	65	70
Pakistan	84	2	2.4	1,024	18	14
Tunisia	77	1	1.3	791	11	13
India	75	1	1.3	14	23	26
Hong Kong	49	0	0.0	33	15	10
Israel	37	0	0.0	350	6	4
Lebanon	23	0	0.0	209	5	7
Egypt	18	0	0.0	17	11	10
Mexico	17	0	0.0	194	3	3
Sudan	11	0	0.0	5	8	9
Taiwan	5	0	0.0	<1	1	1

3.1.3 Simulation Experiment

The simulation experiments are documented in Robinson et al. (2012). Here we present a brief overview of the approach, and results and caveats.

Table 3.4: Cross-tabulation of consignment counts by country and the appearance (TRUE) and non-appearance (FALSE) of the word `FRESH` in the goods description for all countries with at least 5 consignments.

Country	FALSE	TRUE
Bahrain	1	4
China	948	0
Egypt	38	1
Hong Kong	97	0
India	102	0
Iran Islamic Republic Of	1413	1
Israel	56	1
Lebanon	36	0
Mexico	41	1
Pakistan	110	1
Saudi Arabia	109	13
Sudan	12	0
Taiwan	9	0
Tunisia	99	2
Turkey	115	0
United Arab Emirates	382	31
United States	245	79

Introduction

The simulation experiments were undertaken to provide insight into the validity and suitability of different inspection regimes. DAFF has a large volume of interception data, comprising plant pests classified to various taxonomic levels from pathways. These data provide a valuable risk profile for commodities coming from specific origins and therefore a basis for comparing the likely utility of different sample designs.

Knowing the theoretical properties of statistical estimates is useful when nothing is known about the properties of the process. However, if suitable historical inspection data are available, then simulations of the algorithms should be undertaken to replace the theoretical assumptions made during model development. Simulations run using real data will provide the most accurate picture of the validity of the required inspection regime. For example, the theoretical performance of the CSP family can be specified if we assume that the underlying approach rate is always the same. This assumption does not sit comfortably with operational experience, nor with our analyses of the inspection histories of the pathways, which show shifts in approach rates (see, for example, Figure 3.2). By using historical inspection data, we allow for the demonstrated propensity of a pathway to have such shifts in approach rates.

Table 3.5: Summary statistics by supplier for dried and fresh dates. *Count* is the number of consignments imported during the study period, *QF* is the count of consignments with contamination of quarantine interest, *QF %* is the percentage of consignments with contamination of quarantine interest, and *Tonnage* is the total tons of product imported during the study period. The *Countries* and *Importers* columns report the number of countries and importers, respectively, that have engaged with the supplier during the time period. We include only those suppliers with at least 15 recent consignments. The *Countries* column reports the number of countries that each supplier has exported from during the time period.

Supplier	Count	QF	QF %	Tonnage	Countries	Importers
a	119	0	0.0	2,054	1	1
b	105	0	0.0	167	2	5
c	93	1	1.1	1,156	2	4
d	91	1	1.1	1,293	2	4
e	80	0	0.0	1,861	1	7
f	76	0	0.0	1,309	1	1
g	68	0	0.0	176	1	1
h	62	0	0.0	1,477	1	2
i	62	0	0.0	7	1	1
j	59	0	0.0	664	1	3
k	43	0	0.0	966	1	1
l	42	0	0.0	40	1	1
m	41	0	0.0	738	1	3
n	40	0	0.0	583	1	1
o	38	0	0.0	647	1	1
p	37	0	0.0	10	2	5
q	35	0	0.0	889	2	5
r	30	0	0.0	4	1	1
s	30	0	0.0	21	2	1
t	27	0	0.0	304	1	1
u	24	0	0.0	229	1	2
v	24	0	0.0	<1	1	1
w	22	0	0.0	417	1	2
x	20	0	0.0	387	1	2
y	20	0	0.0	19	1	2
z	19	0	0.0	10	1	2
A	18	0	0.0	192	1	3
B	18	0	0.0	7	2	2
C	18	0	0.0	91	1	4
D	17	0	0.0	13	1	4
E	17	0	0.0	262	1	2
F	17	0	0.0	13	1	2
G	16	0	0.0	14	2	2
H	15	0	0.0	32	1	1

Table 3.6: Summary statistics by importer for dried and fresh dates. *Count* is the number of consignments imported during the study period, *QF* is the count of consignments with contamination of quarantine interest, *QF %* is the percentage of consignments with contamination of quarantine interest, and *Tonnage* is the total tons of product imported during the study period. The *Countries* and *Suppliers* columns report the number of countries and suppliers, respectively, that have engaged with the supplier during the time period. We include only those importers with at least 15 recent consignments.

Importer	Count	QF	QF %	Tonnage	Countries	Suppliers
a	199	0	0.0	3,666	3	9
b	154	0	0.0	184	1	3
c	145	0	0.0	3,229	4	9
d	118	0	0.0	1,807	4	6
e	111	2	1.8	1,541	6	9
f	81	0	0.0	1,677	4	6
g	74	0	0.0	987	2	5
h	70	0	0.0	93	1	2
i	55	2	3.6	351	3	9
j	43	0	0.0	692	1	2
k	42	0	0.0	40	1	1
l	41	0	0.0	429	2	5
m	40	0	0.0	38	1	8
n	38	0	0.0	150	2	7
o	37	0	0.0	417	3	6
p	33	0	0.0	87	1	9
q	32	0	0.0	492	4	5
r	31	0	0.0	5	1	2
s	30	0	0.0	21	2	1
t	27	0	0.0	24	1	2
u	26	0	0.0	6	2	1
v	24	0	0.0	403	2	3
w	23	0	0.0	434	1	3
x	23	0	0.0	12	1	4
y	22	0	0.0	9	1	3
z	21	0	0.0	375	2	2
A	20	0	0.0	373	1	4
B	18	0	0.0	25	1	6
C	18	0	0.0	393	2	2
D	18	0	0.0	19	1	4
E	18	0	0.0	68	1	5
F	16	0	0.0	97	1	4
G	15	0	0.0	141	1	1

Design

The simulation experiments are designed as follows. We nominated a range of different inspection designs, as outlined above, and for each selected a range of control parameters.

- CSP–1, CSP–2, and CSP–3: all combinations of monitoring fraction f = 0.1, 0.25, 0.33, and 0.5, and clearance number i = 5, 10, 20, and 40; repeated 100 times for each combination.

We looped through the inspection history of the product in the following way. See Table 3.7 for a row-by-row example of the process.

1. For each line, we determined whether or not the pathway(s) that the line belonged to were flagged for enhanced inspection or monitoring. If any pathway was flagged for enhanced inspection, then the line was inspected. If not, then the line was inspected with probability equal to the monitoring fraction. That is, for each line that was flagged for monitoring, a random number between 0 and 1 was generated, and compared with the nominated monitoring fraction f. If the random number was below f, then the item was 'inspected', otherwise it was released.
2. Simulated detection involved generating another random number between 0 and 1, and comparing this number with the nominated inspection effectiveness e. If the random number was below e *and* the item was contaminated with quarantine risk material, then the contamination was detected. Otherwise the item was passed.
3. The pathways to which the item belonged were then flagged according to the inspection algorithm, and the updated status was used for subsequent inspections.

One of the strengths of the CSP family of algorithms is that it provides an easy way to focus inspection resources on specific parts of the pathway, or sub-pathways. We examined four different strategies for splitting the pathway:

1. no splitting,
2. split by importer,
3. split by supplier, and
4. split by both importer and supplier.

The results at the end of each simulated inspection run were aggregated. Here we report two performance indicators: the simulated post-intervention compliance for the pathway (PIC, Robinson et al., 2011) and the expected number of non-compliant consignments leaked during the time period under scrutiny. A brief summary of the PIC is provided in Appendix A.

Assumptions

The results are based on several important assumptions, listed with a brief commentary as follows.

Table 3.7: Simulation walk-through for CSP–1, using an import compliance clearance number 2 and monitoring fraction 50%. The columns to the left of the vertical bar are sourced from the inspection history. *Cons.* identifies the order in which the consignments arrive. *Fail* reports whether or not the known inspection history for the consignment was a quarantine fail. *Mode* reports whether the pathway is in Enhanced inspection or Monitoring mode at the time of inspection. The x_i and x_d columns report the random numbers that correspond to inspection and detection respectively; if $x_i < 0.5$ then a monitored consignment is inspected, and if $x_d < 0.9$ then a contaminated consignment that is inspected is intercepted. The *Insp.* and *Det.* columns record these outcomes. Finally, *Count* reports the number of inspections since the last detected contamination.

Cons.	Fail	Mode	x_i	x_d	Insp.	Det.	Count	Notes
1	N	E	0.40	0.80	Y	N	1	Enhanced Inspection — clean.
2	N	E	0.90	0.70	Y	N	2	Clean. Move to Monitoring.
3	N	M	0.70	0.50	N	N	2	Not inspected.
4	N	M	0.40	0.80	Y	N	3	Clean.
5	Y	M	0.30	0.20	Y	Y	0	Contaminated. Enhanced mode!
6	N	E	0.90	0.30	Y	N	1	Clean.
7	Y	E	0.90	0.95	Y	N	2	Contamination not detected. Move to monitoring mode.
8	N	M	0.70	0.20	N	N	2	Not inspected
9	N	M	0.80	0.90	N	N	2	Not inspected
10	N	M	0.25	0.55	Y	N	3	Clean.

- Inspection effectiveness is assumed to be 90%. This assumption was made based on discussions with the pathway managers. Leakage surveys that could be used to estimate inspection effectiveness are not available on this pathway. It is possible that the inspection effectiveness may vary by consignment characteristics; for example consignment size. It is certain that effectiveness is depends on within-consignment prevalence.

  This assumption is important because without assuming some value for inspection effectiveness, we would have been unable to estimate the pathway leakage or compliance, and therefore unable to compare the different inspection algorithms. If this assumption of 90% effectiveness is wrong then the estimated leakage rates and compliance rates will be inaccurate.
- The data are assumed to be representative of commercial consignments for each pathway. In fact, we only have access to the data for the entries that were recorded in AIMS. There may be relevant consignments in the Self-Assessed Clearance (SAC) pathway, which do not necessarily appear in AIMS, but these are likely to be small, as the upper limit for declared SAC value is $1000. Also, there may be other consignments of the same product but with incorrect tariff codes.

  This assumption is important because it allows us to draw conclusions about the whole plant product pathway from the available inspection data. If this assumption is wrong then there will be consignments that should be counted that are not. Note that SAC entries that fail inspection are upgraded to AIMS, and therefore will appear in our data.
- We assume for the simulations that the approach rate of contamination in the future will not differ substantially from that of the last two years of the data.

  This assumption is important because it enables us to make some comment about the future performance of the inspection algorithms that we are comparing based on their performance using historical data. If this assumption is wrong then it should rapidly become obvious in the inspection data; the proposed monitoring regime is designed to detect changes in the approach rate with high probability.
- We assume for the simulations that the temporal patterns of contamination in the future will not differ substantially from that of the last two years of the data. That is, we assume that the inspection history represents the full range of patterns of non-compliance that are relevant to the question at hand.

  This assumption is important because our inspection algorithm is designed to respond to patterns of non-compliance, and if future patterns of non-compliance are different from historical patterns then the simulation results will not represent future conditions.

The simulation results are reported only for the period July 1 2008 until the newest entry, although the algorithm was applied to the full dataset (starting at 12 October 2005) as a burn-in strategy. That is, the simulation was performed on the entire inspection history, but we report only the results of the inspections that were undertaken since July 1 2008. This temporal window was selected as providing a reasonable compromise between timeliness, that is, not including any data that were too old, and having enough data available for the comparisons to be valid. See Section 3.1.1.

The analysis can be repeated with more up-to-date data as they become available. The full suite of simulations takes a few hours on a Sun SunFire X4600M2, with eight

AMD Quadcore 2.3GHz CPUs and 64GB memory, running Red Hat Enterprise Linux Server 5.6 (RHEL 5.6).

Results

The results are presented in Figures 3.3 and 3.4. Figure 3.3 shows simulated Post-Intervention Compliance (PIC) against inspection effort for dates inspection history, and is explained in greater detail below. Figure 3.4 shows simulated leakage count against inspection effort for dates inspection history. We note the following points, and emphasize that these points are relevant only to the simulated inspection of dates based on existing data.

Figure 3.3 provides the average simulated post-intervention compliance as a function of inspection strategy and sample size for a range of options. The x-axis is the amount of effort, and less is preferred. The y-axis is the PIC, and higher is preferred. The grey line shows the expected trade-off for random monitoring, and may be used as a baseline to assess the improvement resulting from selecting a CSP strategy.

- The pathway has an inherently low failure rate, so the expected compliance rate (PIC) is higher than 99% in any case.
- All three CSP regimes improve upon random sampling, in that the average PIC returns for each can be substantially higher than the grey line.
- The difference between the different CSP strategies is minimal. However, of the three CSP regimes, CSP–3 seems to provide the best match with the pathway manager's goals.
- The stratification, monitoring fraction, and clearance number all have an effect on the leakage, and the effects can interact with one another.
- Stratifying by importer or supplier or both seems to improve the performance of the algorithm. DAFF has direct interaction with importers so it makes sense to stratify by importer or by importer and supplier.
- Given CSP–3 and stratification by importer or importer and supplier, the best strategy seems to be to set the monitoring fraction to 0.1 and the clearance number to 10.

Figure 3.4 provides leakage count profiles for all the simulations. It can be used as an alternative means for interpreting the above reasoning.

Caveats

In addition to the assumptions outlined above, it is important to note that we have only been able to perform the simulation experiments using the available inspection data. Many of the smaller suppliers will not have had sufficient time to reach the clearance number and move to a sampling regime. Hence if the pathway is largely clear, as we suspect, we can expect that the realized inspection rate will decrease with time, as the smaller importers come online. From this point of view, the results of the simulation experiment are conservative: in time, the same levels of compliance will be able to be maintained with less effort.

On the other hand, we had to assume that (i) the inspection history represents the full range of patterns of non-compliance, which is questionable, and also that (ii) all

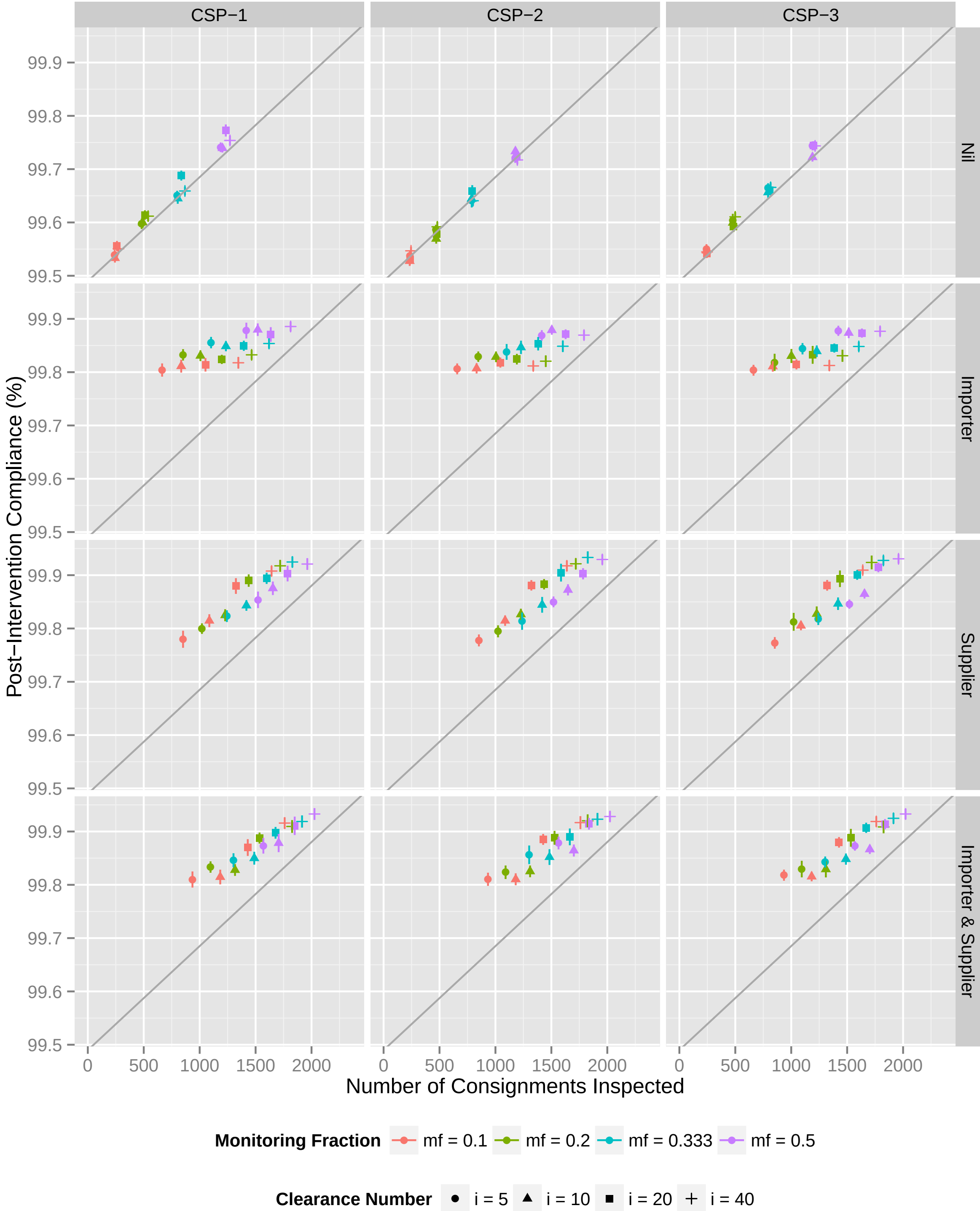

Figure 3.3: Simulated Post-Intervention Compliance (PIC) against inspection effort for dates inspection history. The inspection strategies are in columns, and the stratification options are in rows. Within each panel, the monitoring fraction is delineated by symbol colour, and the clearance number is delineated by the symbol shape. The grey line represents the expected PIC that would result from random sampling. Each symbol has an approximate 95% CI represented as a vertical bar.

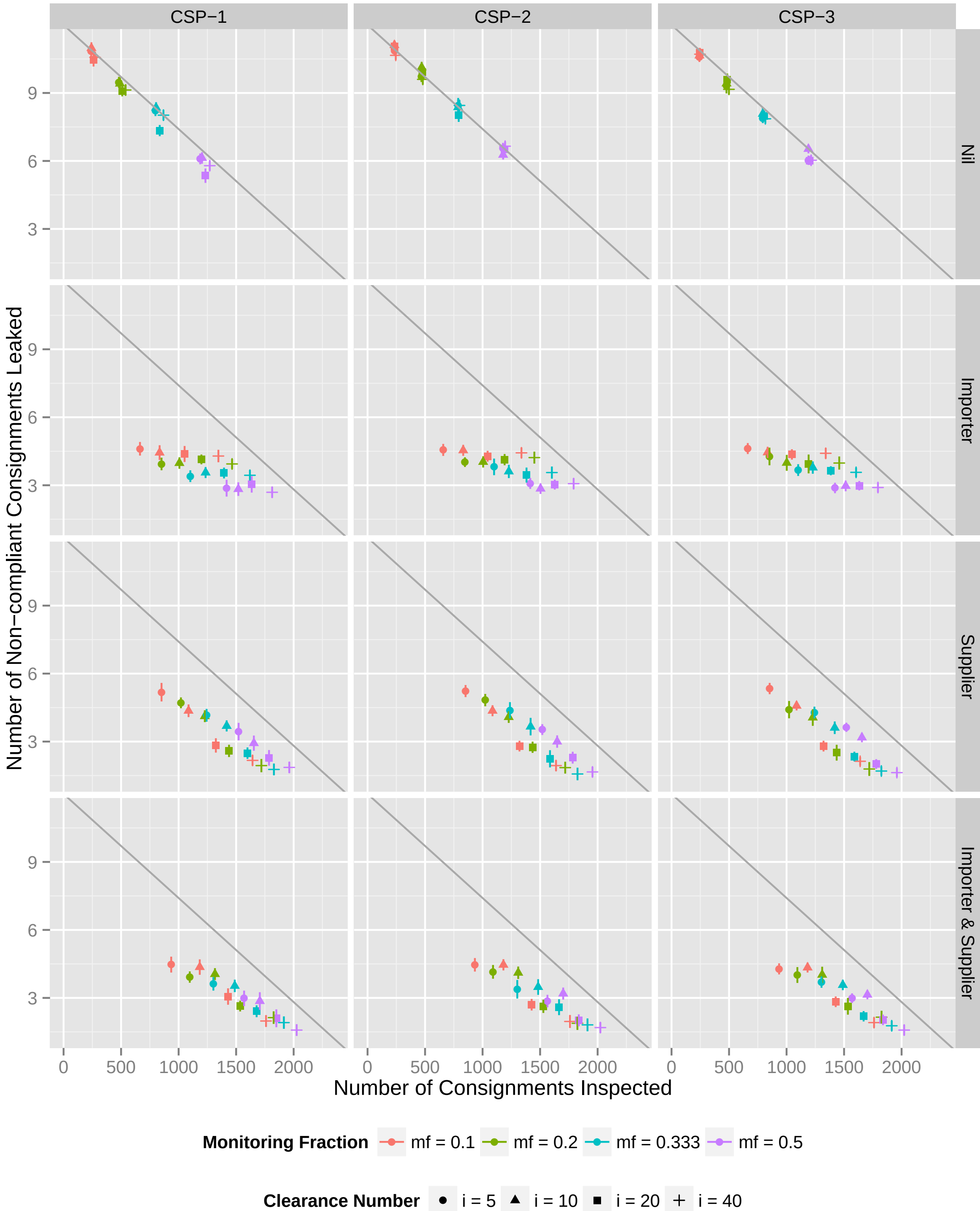

Figure 3.4: Simulated leakage count against inspection effort for dates inspection history. The inspection strategies are in columns, and the stratification options are in rows. Within each panel, the monitoring fraction is delineated by symbol colour, and the clearance number is delineated by the symbol shape. The grey line represents the expected leakage that would result from random sampling. Each symbol has an approximate 95% CI represented as a vertical bar.

the non-compliance that is relevant to our experiment has been detected, which is also questionable. That is, the true failure rate will probably be marginally higher than is represented here. From these points of view, the results of the simulation experiment are anti-conservative: given the amount of effort invested, the true resulting compliance will probably be less than is forecast here.

Implementation

CSP-3 was implemented for the fresh and dried dates pathway in July 2012, with monitoring fraction 0.1 and clearance number 10.

3.1.4 Inspection Regime

The recommendations that arise from this exercise are predicated on two things. First, zero risk is unachievable and undesirable citepbeale-2008, so the risk has to be balanced against the cost of risk management. Second, it is very difficult for DAFF to determine a complete framework in which the costs of mitigation of the risks of invasion are balanced against the potential effects of invasion (although such work is underway within DAFF). Therefore a decision must be made as to how much effort is spent on mitigating risks on a pathway, and pathways that are deemed low risk, because they have very little contamination and what contamination they have is deemed to not be of biosecurity concern, will need to have less intervention. The CSP algorithm has been introduced for managing such low-risk pathways (Robinson et al., 2012). Until such time as a complete risk-based framework has been developed, the nomination of monitoring goals for CSP must be undertaken based on the judgment of the pathway managers where policy does not prescribe an intervention level.

Recommendation: CSP–3

We recommend the deployment of a continuous sampling plan, CSP–3, on a importer-specific basis for this import pathway.

The basis of this recommendation is the outcome of a large number of simulation experiments performed using 5.5 years of historical inspection data from 2005–2012 for dates. It is important to note that the simulation experiments provided only modest evidence for the superiority of CSP–3 over alternatives such as CSP–1. However, the interpretation of the results of the simulation experiment has to be tempered by a recognition that the historical inspection data cover only a few years of interaction.

The reason that CSP–3 is chosen ahead of CSP–1 is that it allows for the possibility of isolated leakage incidents, or random once-off non-compliance, without shifting immediately to a census mode and penalizing importers with 100% inspection rates. The reason that it is chosen ahead of CSP–2 (not reviewed here) is that it provides temporary increased scrutiny of the pathway to see if a leakage incident seems likely to be part of a trend, and not simply an isolated failure. Therefore CSP–3 represents a compromise between the measured approach of CSP–2, which allows for some leakage, and the focused approach of CSP–1, which keeps watch for upward shifts of the failure rate.

The reason that we recommend that CSP–3 be deployed by importer, as opposed to by supplier, or both, is partially due to the results of the simulation experiment, and partially due to common sense. The results of the simulation experiment showed reasonable support for an importer-specific approach. Also, we think that it is likely and preferable that if a quarantine failure is observed then the importer of the consignment should receive greater scrutiny. Therefore the recommendation is based on a combination of simulation results, prudence, and knowledge about the pathway.

For CSP–3, the inspection algorithm is as follows,

- M is satisfied by the observation of i successive compliant consignments.
- C is satisfied by two observations of non-compliance that are within k inspections of one another. Usually, $k = i$.

Note that *in addition to the above prescription*, when any non-compliance is detected, the next four consignments will all be inspected, i.e. the monitoring process will be suspended for the next four consignments. This is a temporary shift and is done regardless of whether the non-compliance is the first or second detected in any number of units.

To summarize CSP–3, if the system is in monitoring mode, then the consignments are being randomly selected for inspection at rate f. If a non-compliant consignment is intercepted, then the next four consignments are inspected, and if they are all compliant, the system returns to monitoring. If another non-compliant consignment is found within i inspections of the previous non-compliance, then the system switches to census mode.

The entire algorithm is specified in terms of just two parameters: i and f. The trade-off between the parameters i and f can be interpreted as follows. A high i means that once a pathway is in census mode, it will take a larger number of compliant consignments in order to achieve monitoring mode. Consequently, we can think of i as being the amount of evidence that we need for concluding that a non-compliant importer has cleaned up their pathway, or a penalty upon the importer for a non-conformity. A high f means that we want to be more likely to detect changes in the underlying rate quickly.

CSP–3 Parameters We used a set of simulation experiments to develop a recommendation for values to use for the monitoring fraction f and the clearance number i. These experiments are documented in Section 3.1.3. Based on the results, the following set of parameters seem tentatively suitable.

- $f = 0.1$; $i = 10$ (post-intervention compliance $> 99.8\%$; total inspection count of just over 1000 out of 2349 over two and a half years).

3.1.5 Recommendations

We recommend the deployment of *CSP–3* for the dried & fresh dates pathway. Briefly, the inspection regime will adopt the following pattern, for each *importer*:

1. inspect all items in a census mode, until 10 (for example) successive compliant consignments have been observed. Then,

2. switch to monitoring mode, inspecting at rate 10% until a quarantine non-compliance is detected. Following a non-compliance, the next four consecutive consignments must be inspected, after which monitoring is again used.

   If a second non-compliance occurs within 10 (for example) inspections of the previous non-compliance, then switch back to the above census mode.

The selection of clearance number 10 and monitoring fraction 10% are based on simulation experiments that were conducted using historical dates inspection data.

Based on the simulation experiment, we expect that this strategy would have resulted in a pathway post-intervention compliance rate $> 99.8\%$, using a total inspection count of just over 1000 out of 2349 over the last two and a half years.

3.1.6 Outcome

As a result of this analysis, the risk-based approach to inspection for dates (dried and fresh) was implemented starting July 16 2012.

3.2 Medium-Risk Nursery Stock

3.2.1 Introduction

The purposes of this case study were three-fold. First, the pathway manager wanted to assess the evidence for there being variation in the quarantine interception rates across the regions. Second, the pathway manager wanted to know what was the effect of altering the definition of quarantine failure from any contaminated plants to an upper limit, for example, greater than two contaminated plants in the consignment might result in a quarantine detection. Third, the pathway manager wanted to establish the effects on efficiency and effectiveness of altering the inspection protocol from the current inspection of all the plants in the consignment, which can number up to the hundreds of thousands, to a sample, for example the 600 units favoured in many other pathways. We removed this third request during the project after discussion with the DAFF project manager.

The discussion on risk and risk mitigation on page 25 is also relevant here.

Import Conditions

Imported nursery stock is living plant material with the capacity to introduce exotic plant pests and diseases into Australia. Medium risk nursery stock (MRNS) is any live plant (excluding tissue cultures) that has not met the high risk classification[1]. MRNS have a potential to harbour quarantine pests, however, due to the significant number of species assessed for import each year, a detailed Pest Risk Analysis (PRA) is not carried out on each MRNS species. MRNS are required to undergo a period of post-entry quarantine (PEQ) in a Quarantine Approved Premise (QAP) to mitigate possible quarantine risks and maintain Australia's Appropriate Level of Protection (ALOP), which aims to reduce risk to a very low level.

Current MRNS policy requires all plants in a consignment to be subjected to a 100% visual inspection on arrival, followed by a minimum three-month period of growth within a post-entry quarantine facility and two subsequent DAFF inspections performed in this period.

Data Preparation

The data were gathered from regional import spreadsheets. During a review of MRNS imports undertaken in 2010 it was determined that the AIMS import system did not provide a level of consignment detail required to make risk based decisions on MRNS import pathways. Monthly consignment reporting sheets were established in January 2011 to capture specific import volume and consignment compliance information for each MRNS consignment, especially the number of plants and the number of noncompliant plants in each consignment.

Analytic Timespan

The analysis was constrained to inspections from January 2011 to December 2011 (totaling 1249 consignments), in order that the results reflect the current status of the pathway as much as possible. Some odd earlier dates were included in the dataset, and should be scrutinized.

[1]FIXME How is this defined?

Pathway Summary

The pathway is summarized in Figure 3.5.

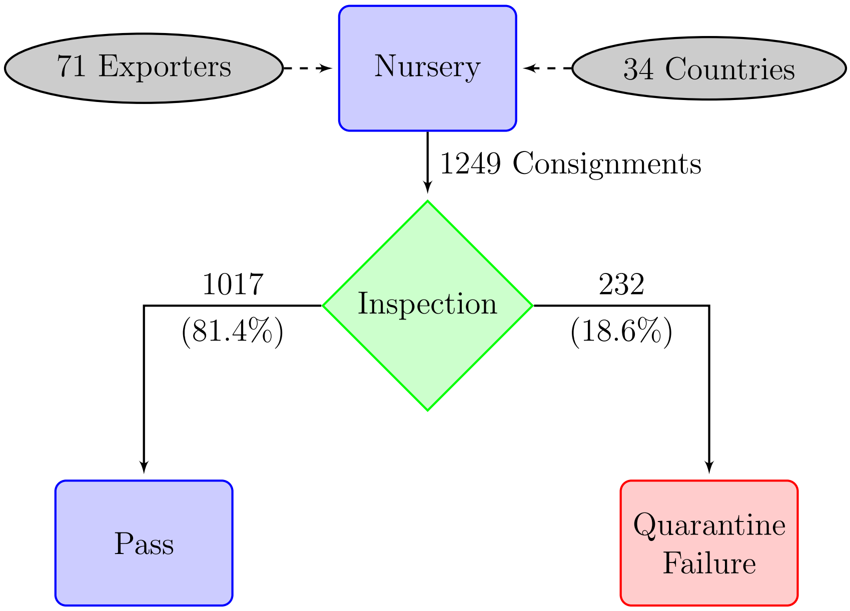

Figure 3.5: Nursery consignments flow chart with statistics for January 2011 to December 2011. A quarantine failure was recorded for consignments with a detection of quarantine concern, such as insect, pathogen, or contamination.

3.2.2 Analysis

Summaries

This section provides a statistical overview of the data. The full dataset comprises 1249 consignments with record creation dates ranging from January 2011 to December 2011, and comprises entries from 34 countries and 96 importers.

For this analysis each individual species line listed within an AIMS entry was considered a consignment. The inspection data should be treated with some caution as four consignments include more than 100,000 individual plants each, and the logistics of thorough inspection of consignments of such sizes are impressive.

A plot of a smoothed average of the consignment-level failure rates against time is presented in Figure 3.6. The figure shows a failure rate increasing from below 15% to around 30%. The consignment-level failure rate for the entire period was 18.57%.

A smoothed plot of the plant failure rate against time is presented in Figure 3.7. The figure shows an average failure rate increasing from below 5% to around 15%. The plant-level failure rate for the entire period was 2.72%.

The following five tables describe the basic characteristics of the pathway. The statistics in Table 3.8 summarize the inspection data for those countries with at least twelve consignments during the key time period. The most substantial exporting country in terms of consignments is the USA, and in terms of plants is China, followed by Indonesia. The plant-level failures rates are particularly high for the USA, Thailand, and Japan, and consignment-level failure rates are also high for the Netherlands, Spain, Vietnam, and El Salvador, although the consignment count for the latter three is much smaller.

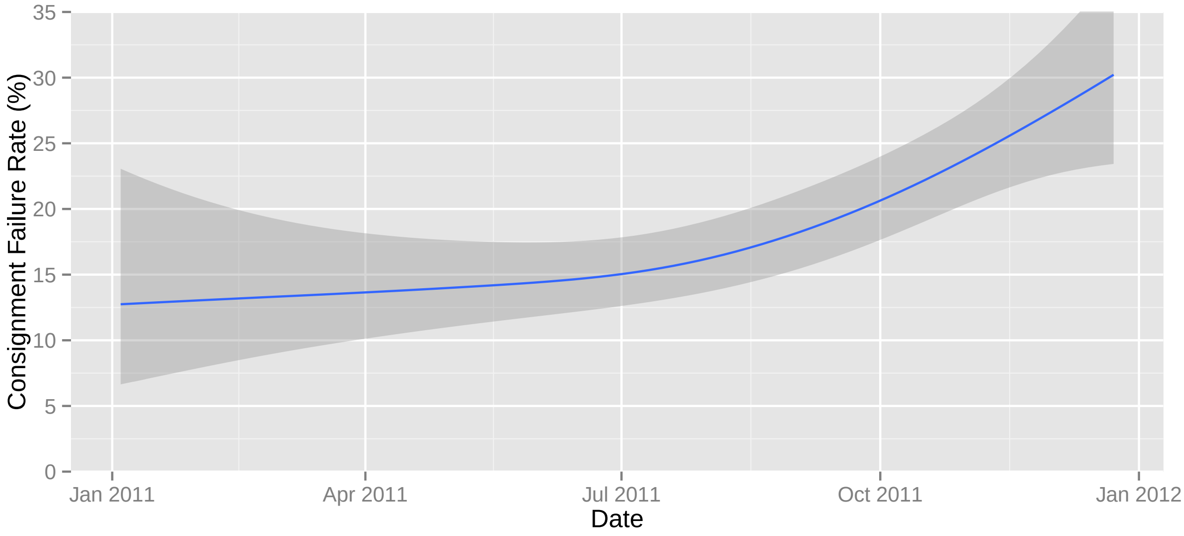

Figure 3.6: Consignment failure rates (%) for nursery stock smoothed by date, with a 95% confidence interval (shaded region) added. The width of the shaded region indicates the uncertainty of the line, which becomes narrower as the sample size increases.

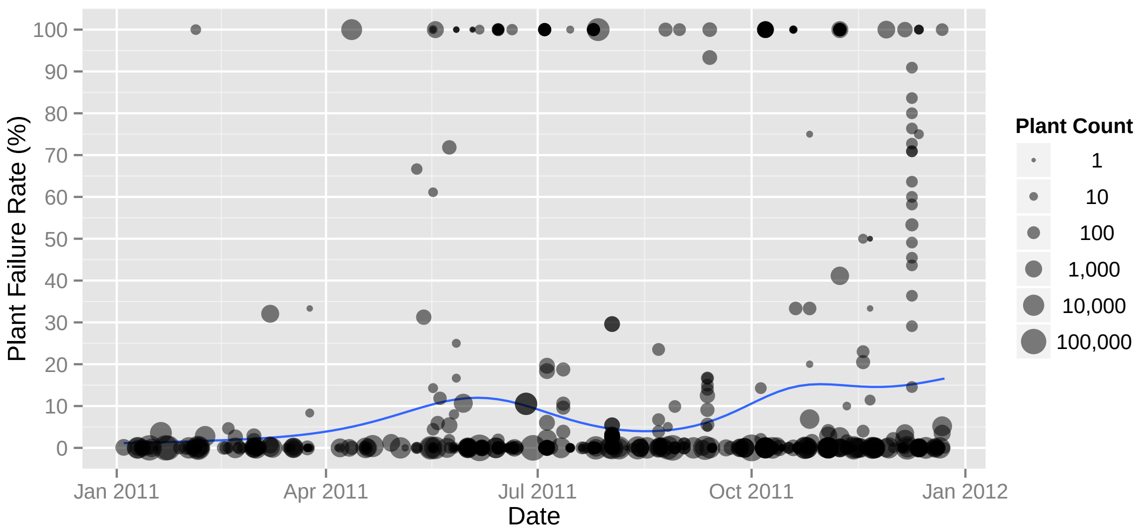

Figure 3.7: Plant failure rates (%) for nursery stock smoothed by date, and not weighted by consignment size. Each point represents a consignment, and the consignment size dictates the point size.

Table 3.9 summarizes the inspection data for exporters with at least twelve consignments. This table shows that information is not available on a number of significant exporters. Profiling by exporter, while superficially attractive, will require more rigorous data collection.

Table 3.10 summarizes the inspection data for importers with at least twelve consignments. In terms of consignments, there is a single large importer, which by volume is 1.5 times the size of the next largest; otherwise there is not much differentiation in the population. The largest importer by consignment has a good inspection record at the consignment level. The largest by volume has an impressive inspection record. Four importers import the bulk of the plants.

Table 3.11 summarizes the inspection data by genus only for those with at least twelve consignments during the key time period. One genus stands out for plant count: Dracaena. A number have very high consignment failure rates, specifically Perlargonium, Rosa, and Hibiscus. Allium and Perlargonium dominate the plant-level failure rates.

Table 3.12 provides the same information for regions. Central East accepts the most consignments, and South East the most plants. Consignment and plant failure rates are high at Central East and North East.

Table 3.8: Summary statistics by country for nursery stock. *Cons* is the number of consignments, by which the table is ordered. *CF* is the number of consignments with any number of contaminated plants. *CP* is the number of consignments with partial failures; that is, full inspection would be needed to guarantee detection. *CR %* is the consignment-level failure rate. *Plants* is the total number of plants imported. *PF* is the count of plants with quarantine failure. *PR %* is the plant-level failure rate. *Reg.* reports the number of regions to which the country's consignments are sent. The *Gen.* column reports the number of genera imported from that country. The *Exp* and *Imp* columns report the number of exporters and importers that have exported from each country during the time period.

Country	Cons	CF	CP	CR %	Plants	PF	PR %	Reg.	Gen.	Exp.	Imp.
USA	493	95	36	19.3	21,864	9304	42.6	4	70	20	28
Singapore	128	2	2	1.6	509	7	1.4	4	59	3	5
United Kingdom	105	5	2	4.8	1,201	6	0.5	2	66	3	7
Japan	66	47	25	71.2	14,230	3453	24.3	4	13	4	5
Netherlands	49	14	14	28.6	104,021	1701	1.6	3	18	4	8
China	42	5	4	11.9	1,142,295	1882	0.2	4	16	11	12
France	41	0	0	0.0	923	0	0.0	3	8	2	4
Indonesia	41	1	0	2.4	257,732	8600	3.3	1	7	6	2
Thailand	31	7	5	22.6	90,600	25640	28.3	3	15	7	9
Israel	25	2	2	8.0	1,217	10	0.8	1	21	2	1
Guatemala	24	2	2	8.3	86,609	66	0.1	2	4	1	3
Germany	21	2	2	9.5	17,563	40	0.2	4	16	2	6
Philippines	21	0	0	0.0	29,300	0	0.0	1	6	1	1
Denmark	19	0	0	0.0	1,380	0	0.0	2	3	2	2
India	17	1	1	5.9	21,480	5	0.0	2	7	1	2
Sri Lanka	17	1	1	5.9	15,068	523	3.5	3	3	3	4
Spain	16	4	2	25.0	11,176	40	0.4	3	8	2	2
Vietnam	16	16	16	100.0	8,024	470	5.9	1	1	1	1
El Salvador	14	12	12	85.7	1,660	218	13.1	1	8	2	1

Table 3.9: Summary statistics by exporter for nursery stock. See Table 3.8 for explanations of the columns.

Importer	Cons	CF	CP	CR %	Plants	PF	PR %	Exp.	Reg.	Gen.	Cys.
a	470	46	36	9.8	1,699,726	40660	2.4	62	5	205	28
b	272	1	0	0.4	764	10	1.3	4	2	35	3
c	39	0	0	0.0	156	0	0.0	2	1	1	1
d	34	0	0	0.0	590	0	0.0	1	1	1	1
e	34	34	17	100.0	2,000	1584	79.2	1	1	1	1
f	33	16	7	48.5	3,460	1024	29.6	1	1	2	1
g	32	0	0	0.0	36	0	0.0	1	1	9	1
h	30	1	0	3.3	50	1	2.0	1	1	2	1
i	25	25	1	100.0	3,083	2675	86.8	1	1	1	1
j	17	0	0	0.0	905	0	0.0	1	1	1	1
k	16	16	16	100.0	624	3	0.5	1	1	16	1
l	16	2	2	12.5	288	4	1.4	1	1	7	1
m	16	6	3	37.5	1,483	73	4.9	1	1	2	1
n	16	16	16	100.0	8,024	470	5.9	1	1	1	1
o	14	8	8	57.1	33,900	1510	4.5	2	2	1	1
p	14	0	0	0.0	201	0	0.0	1	1	1	1
q	14	0	0	0.0	84,862	0	0.0	1	1	6	1
r	13	6	0	46.2	1,350	651	48.2	1	1	2	1

Table 3.10: Summary statistics by importer for nursery stock. See Table 3.8 for explanations of the columns.

Importer	Cons	CF	CP	CR %	Plants	PF	PR %	Reg.	Gen.	Exp.	Cys.
a	179	1	0	0.6	506	7	1.4	1	7	1	1
b	117	62	19	53.0	12,205	4679	38.3	1	20	7	5
c	104	2	2	1.9	1,035,197	66	0.0	1	22	1	5
d	91	0	0	0.0	151	3	2.0	1	27	1	1
e	91	5	2	5.5	683	6	0.9	1	57	1	2
f	69	18	16	26.1	53,708	473	0.9	1	36	9	9
g	39	31	30	79.5	40,125	2103	5.2	1	9	5	3
h	38	35	17	92.1	3,748	2578	68.8	1	2	3	3
i	36	0	0	0.0	171,403	0	0.0	1	11	8	4
j	34	0	0	0.0	590	0	0.0	1	1	1	1
k	32	0	0	0.0	36	0	0.0	1	9	1	1
l	30	0	0	0.0	99	0	0.0	1	30	1	1
m	30	1	0	3.3	50	1	2.0	1	2	1	1
n	28	0	0	0.0	310,736	0	0.0	1	11	1	4
o	28	3	1	10.7	258	40	15.5	2	17	1	2
p	27	1	1	3.7	1,388	30	2.2	1	5	3	3
q	23	2	1	8.7	258,810	8600	3.3	1	6	10	4
r	20	0	0	0.0	80	0	0.0	1	1	1	1
s	19	0	0	0.0	76	0	0.0	1	1	1	1
t	18	9	8	50.0	83,583	7160	8.6	1	14	4	5
u	16	16	16	100.0	624	3	0.5	1	16	1	1
v	16	2	2	12.5	288	4	1.4	1	7	1	1
w	14	0	0	0.0	201	0	0.0	1	1	1	1

Table 3.11: Summary statistics by genus for nursery stock. See Table 3.8 for explanations of the columns.

Genus	Cons	CF	CP	CR %	Plants	PF	PR %	Cys.	Imp.	Reg.	Exp.
Agave	61	1	1	1.6	159	2	1.3	1	3	1	2
Allium	45	6	0	13.3	4,716	4560	96.7	1	3	1	2
Rosa	44	38	20	86.4	13,218	3331	25.2	5	6	3	2
Haworthia	41	0	0	0.0	5,168	0	0.0	2	3	2	2
Argyranthemum	35	6	3	17.1	2,513	73	2.9	4	3	2	5
Chrysanthemum	35	8	8	22.9	6,205	49	0.8	4	6	4	3
Vriesea	35	2	2	5.7	2,150	2	0.1	3	4	2	3
Dendranthema	34	0	0	0.0	590	0	0.0	1	1	1	1
Dracaena	33	3	2	9.1	1,146,734	9405	0.8	4	5	3	6
Bulbophyllum	28	0	0	0.0	37	0	0.0	1	1	1	1
Hibiscus	28	17	17	60.7	8,615	471	5.5	3	4	2	3
Phragmipedium	28	1	0	3.6	55	1	1.8	2	2	2	2
Perlargonium	25	25	1	100.0	3,083	2675	86.8	1	1	1	1
Crassula	23	1	0	4.3	68	7	10.3	1	1	1	1
Discorea	23	0	0	0.0	44	0	0.0	1	1	1	1
Echeveria	23	0	0	0.0	67	0	0.0	1	1	1	1
Phalaenopsis	23	1	0	4.3	100,689	442	0.4	4	5	4	8
Tillandsia	22	2	2	9.1	70,708	66	0.1	2	2	2	2
Anthurium	18	9	9	50.0	45,016	1510	3.4	2	4	4	2
Aster	17	10	0	58.8	1,753	1022	58.3	2	2	2	2
Paphiopedilum	16	2	1	12.5	213	96	45.1	2	4	2	4
Nepenthes	15	0	0	0.0	437	0	0.0	1	2	1	2
Dendrobium	13	2	1	15.4	80,559	24755	30.7	3	4	3	4
Osteospermum	13	1	1	7.7	1,451	20	1.4	5	5	3	5

Table 3.12: Summary statistics by region for nursery stock. See Table 3.8 for explanations of the columns.

Region	Cons	CF	CP	CR %	Plants	PF	PR %	Cys.	Gen.	Imp.	Exp.
Central east	694	161	78	23.2	87,281	14687	16.8	20	98	31	26
South east	277	8	5	2.9	1,366,975	96	0.0	18	105	21	1
North east	148	57	50	38.5	177,597	33771	19.0	20	87	17	22
South west	99	6	5	6.1	471,395	8631	1.8	17	31	27	31
Northern	30	0	0	0.0	99	0	0.0	1	30	1	1

3.2.3 Regional Analysis

We begin the analysis with a set of unformatted consignment-level cross-tabulations that show clearly that there is a pattern of dependence between receiving region and supplying country, genus, importer, and exporter. We then fit a statistical model and provide estimates of region-level detection rates, and confidence intervals, to enable an informal comparison.

Cross-tabulations

For all the cross-tabulations in this section we include only those levels that have 12 or more consignments in the time period, in order to eliminate distracting clutter.

Country

```
                Region
Country           CE  NE   N  SE  SW
  China            1   4   0  26  11
  Denmark         17   0   0   2   0
  El Salvador      0  14   0   0   0
  France          35   0   0   5   1
  Germany          5  13   0   1   2
  Guatemala        1   0   0  23   0
  India            1   0   0  16   0
  Indonesia        0   0   0   0  41
  Israel           0  25   0   0   0
  Japan           60   4   0   1   1
  Netherlands     21  15   0   0  13
  Philippines      0   0   0  21   0
  Singapore       91   6  30   1   0
  Spain            7   1   0   8   0
  Sri Lanka       15   1   0   0   1
  Thailand         2  22   0   0   7
  USA            409  26   0  54   4
  United Kingdom   0   0   0 100   5
  Vietnam         16   0   0   0   0
```

Genus

Genus	Region CE	NE	N	SE	SW
Agave	61	0	0	0	0
Allium	45	0	0	0	0
Anthurium	13	1	1	0	3
Argyranthemum	33	2	0	0	0
Aster	16	1	0	0	0
Bulbophyllum	28	0	0	0	0
Chrysanthemum	24	1	0	1	9
Crassula	23	0	0	0	0
Dendranthema	34	0	0	0	0
Dendrobium	7	2	0	0	4
Discorea	23	0	0	0	0
Dracaena	0	2	0	9	22
Echeveria	23	0	0	0	0
Haworthia	40	0	0	1	0
Hibiscus	27	0	0	1	0
Nepenthes	15	0	0	0	0
Osteospermum	8	3	0	2	0
Paphiopedilum	15	1	0	0	0
Perlargonium	25	0	0	0	0
Phalaenopsis	7	1	0	3	12
Phragmipedium	27	0	0	0	1
Rosa	36	2	0	6	0
Tillandsia	0	1	0	21	0
Vriesea	0	3	0	32	0

Importer

Importer	Region CE	NE	N	SE	SW
Adrian Antonello	19	0	0	0	0
Alan Carle	0	0	30	0	0
Benara	0	0	0	0	36
Dans Plants	0	0	0	28	0
David Lewis	20	0	0	0	0
Eureka Plants	34	0	0	0	0
Flora International P/L	38	0	0	0	0
Ian Chalmers	91	0	0	0	0
Lambley Nursery	7	0	0	21	0
Malcolm Cameron	0	16	0	0	0
Nola Carr	179	0	0	0	0
Oasis Horticulture P/L	117	0	0	0	0
Paradisia P/L	0	0	0	104	0
Pearce's Nurseries	0	18	0	0	0
Peter Hopkinson	30	0	0	0	0
Plant Breeding Institute	27	0	0	0	0
Plant Growers Australa P/L	0	0	0	91	0
Plantation 2000	0	16	0	0	0
Propagation Australia P/L	0	69	0	0	0
Rebecca Lambkin	32	0	0	0	0
Sprint Horticulture P/L	39	0	0	0	0
Tropical Colours	0	0	0	0	23
Victor Franco	14	0	0	0	0

Exporter

Exporter	Region CE	NE	N	SE	SW
Anthura	13	0	0	0	1
Ball Horticultural Company	25	0	0	0	0
Borneo Exotics	14	0	0	0	0
Boundary Garlic	39	0	0	0	0
Challet-Herault SAS	34	0	0	0	0
Evelyn Welbaum	32	0	0	0	0
Flori Partner A/S growing solutions	17	0	0	0	0
J and H Japan Inc	34	0	0	0	0
Michael's Bromeliads INC	0	16	0	0	0
PT Benar Flora Utama	0	0	0	0	14
Piping Rocks Orchids	30	0	0	0	0
Plantation 2000	0	16	0	0	0
Sakata	13	0	0	0	0
Suntory Flowers	16	0	0	0	0
Syngenta Flowers	33	0	0	0	0
Various	270	0	0	0	2
Young Plants	16	0	0	0	0
unknown	60	59	30	277	44

3.2.4 Testing the Regional Effect Using a Statistical Model

Our next goal is to determine the strength of evidence that the rate of failures at one region or other differs once all other factors have been taken into account.

Modelling Approach

The usual statistical approach to detecting patterns in the failure rates of different categories when the outcome is binary (pass/fail) is to fit a *generalized linear model* to the data, and to try to assess whether there is any substantial difference (defined in a statistical way) between the failure rates of the categories. See, for example, McCullagh and Nelder (1989) or Hilbe (2009).

The usual model-fitting approach leads to some problems with these data. First, it doesn't allow for the effects of other potential predictor variables, such as *country of origin*, *exporter*, and *genus*, unless those are included in the model. But then the problem of developing a suitable summary remains. Second, there are no fails at all in the *North* region, which means that usual models will estimate the failure rate for that region as zero, which seems unrealistic.

We solved these problems as follows. First, we fit a Bayesian logistic model to the data which will perform some shrinkage of predictions, using the *arm* package in R (Gelman et al., 2011)[2]. We used all the candidate predictor variables for this model fit, namely, country, exporter, importer, genus, and then region. We then developed parameter estimates and confidence intervals using the `estimable` function from the `gmodels` package (Warnes, 2011a). We converted the estimates and intervals to probabilities, to facilitate comparison and interpretation, and pro-rated those estimates and intervals so that the average predicted probability aligns with the observed failure rate. We created the plot presented in Figure 3.8 using the `plotCI` function from the `gplots` package (Warnes, 2011b).

In order to be confident that the results were reasonably robust to our model assumptions we tried several other modelling approaches, and none suggested a different conclusion.

Without going into too much detail, even after taking account of different importing countries, exporters, importers, and genera, there does seem to be statistical evidence that the failure rate differs among regions. Figure 3.8 summarizes the outcome, showing that CE and NE have unusually high failure rates, or equivalently that N, SE, and SW have unusually low failure rates.

3.2.5 Changing the Failure Definition

It was also of interest for the pathway manager to determine what was the effect upon the approach rate of the specific definition of failure at the consignment level, namely, that a consignment holding any contaminated plants would be considered actionable. That is, the manager wanted to know whether a majority of the consignments that failed did so because they had only very small numbers of contaminated plants. Figure 3.9 provides a summary of the effect upon the estimated approach rate of increasing the cutoff of the number of allowable contaminated plants from zero. It can be interpreted in the following

[2]This particular model scales all the predictors and then uses a Cauchy prior on the parameter estimates; see Gelman et al. (2008).

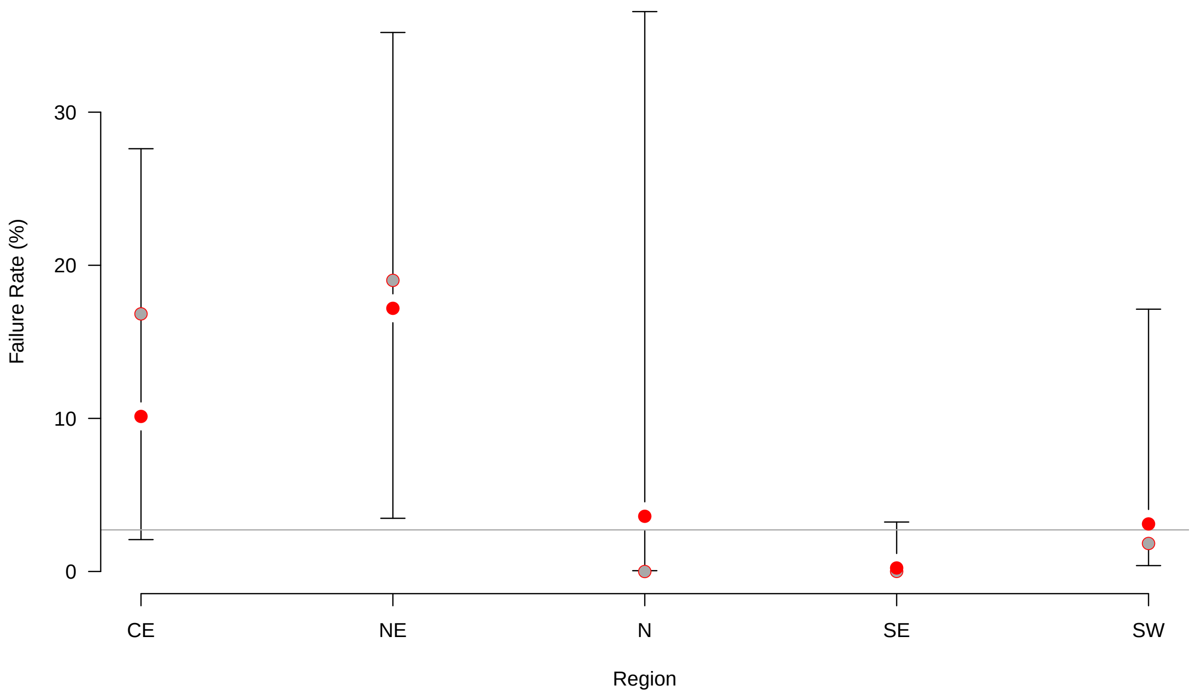

Figure 3.8: Statistical comparison of regional failure rates. The red dot is the estimated regional failure rate after allowing for different patterns of importing country, exporter, and genera. The grey dot is the "raw" failure rate for the region. The grey bar provides a 95% confidence interval for the regional failure rate.

way. As the failure condition increases, that is, as more contaminated plants are allowed before a consignment is declared actionable, there will be fewer actionable consignments.

The hope was that there might be a relatively sharp drop at some point, but this hope was not realized. For example, the approach rate based on the current definition (no tolerance) is 18.5%. If the tolerance is increased to 10 contaminated plants then the approach rate will be slightly under 12.5%. If the tolerance is increased to 20 contaminated plants then the approach rate will be slightly under 10%. We do not see strong evidence here that increasing the tolerance to contamination in the consignments will change the interpretation of the resulting pathway statistics.

In any case, any shift from zero to non-zero tolerance for contamination within consignments would require assessment of the concomitant increase in biosecurity risk, which was beyond the scope of this project. Without a formal assessment the precautionary principle should continue to apply in the management of biosecurity risk in this pathway.

3.3 Plant-Based Stockfeed

Although not in the original prospectus, it was deemed useful to attempt a similar analysis of the plant-based stockfeed (PBS) pathway. We attempted to obtain a suitable database that would permit assessment of the PBS pathway for a risk-based approach. This section summarizes our failure to progress the analysis in this regard.

Analysis of the PBS pathway has been complicated by operational constraints. The primary challenge has been to extract from AIMS just those quarantine entries that correspond to PBS consignments. This is a challenge because there is no unique tariff

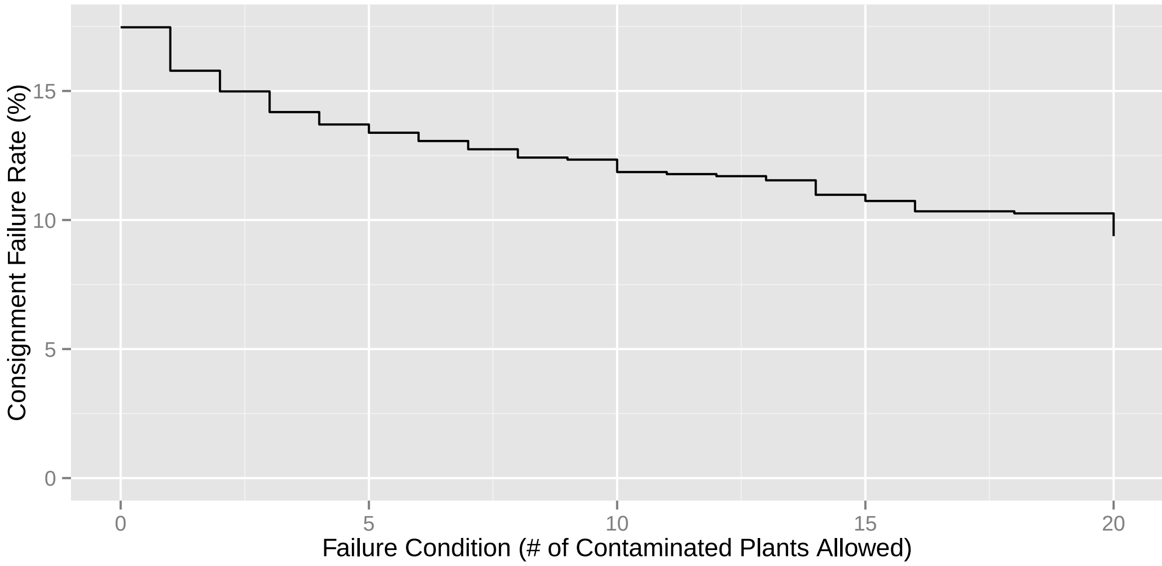

Figure 3.9: Consignment failure rates (%) against failure count cutoff for nursery stock. Each point represents the failure rate that would correspond to the cutoff represented on the x-axis.

code, or set of tariff codes, that clearly distinguishes PBS, as tariff codes are constructed for economic purposes, not for biosecurity.

We began with a relatively modestly scoped data dump from AIMS that captured all the tariff codes that were most likely to correspond with PBS. After cleaning, this database was too small, and importantly, omitted some entries that were known to the authors on the basis of being infamous quarantine fails.

We then tried an AIMS dataset that contained all the tariff codes that might contain PBS. This dataset comprised

- 640,219 lines,
- 38,484 quarantine entries,
- 2293 importers (2419 importer codes, n.b. 2419 > 2293),
- 2431 suppliers (2587 supplier codes, n.b. 2587 > 2431)
- 53 tariff codes,
- 934 permit codes (stored in the Import Permit field), and
- 8084 goods descriptions.

We next tried to winnow this database down to the entries that were of interest.

3.3.1 Filtering by Permit was Unsatisfactory

After considerable discussion, we tried the following strategy. The one unique characteristic for PBS consignments is that importers require a permit for importing from each supplier. Permits for bulk pathways are granted by DAFF staff after an audit of the supplier, and last for two years. The entire collection of permits that has been granted by DAFF is available as a spreadsheet. We filtered this spreadsheet using the names of the

DAFF staff who handled PBS permitting for the last eight years. This provided a list of the permit numbers, importer names, and supplier names.

In theory, the permit number should be recorded against the AIMS quarantine entry. We tried to filter the AIMS entries to only those that had permit numbers that appeared in the permits database. Permit numbers typically comprise a 4-digit year followed by a 5-digit number, but sometimes seven digits preceded by IP. Also, import permits may appear in the Import Permit field (934 unique), the Permit Comment field (11191 unique). There is also a permit comment 2 field which seems to largely mimic the permit comment field in content.

This filtering was impossible without considerable manual work because the format of the permit recording was highly variable; for example, often several import permits are recorded against the same quarantine entry (presumably because the entry contains numerous lines) and sometimes in a truncated format. Examples follow.

```
IP10013015
IP09016710, IP09016716
IP09003084 AND IP09000460
IP10018184-IP10018187
IP11000654,656,658,659
IP11000654,656,658,659,IP11001746
IP11000659,58,56,54,1100746
IP11006531,6533-6535
IP10012265,12288,12261,12263
IP10012265/12263/12288/12261
IP09008926 REPLACED BY IP09013144
```

Some of these combinations would be relatively easy to sort out using text-manipulation tools, but others, particularly the truncated records, would take much more work.

The other significant problem would be in determining exactly which lines the permits correspond to. We remark that many quarantine entries are recorded in *container mode*, as opposed to *line mode*. This means that the same information is recorded against all the lines in the entry, including intervention information such as direction, direction category, direction comments, standard comments, field comments, and so on. The recording of quarantine entries in container mode is probably the most significant impediment to strategic use of quarantine inspection data for risk-based management.

We tried to use a filter that was based on a straight match of the permit numbers from the AIMS entry against the permit numbers as recorded in the DAFF permit database. This filtering left far too few AIMS entries.

3.3.2 Filtering by Supplier and Importer was Unsatisfactory

We next tried to identify the relevant consignments using the combinations of supplier and importer that appeared in individual permit applications. We reasoned that it was likely, although not guaranteed, that any consignments between such a matched pair during the period for which a PBS import permit was in place would be for PBS.

The problem with this solution was matching up the importer and the supplier from the permit database. The permit database includes the importer and supplier names as free text entries, but not the importer or supplier codes. We tried to infer importer and supplier codes for importer and supplier names respectively, using AIMS as a source. This was problematic because a number of supplier names corresponded to more than one code, indeed one name had seven codes. A similar pattern occurred with importer names.

A further problem is that only 137 out of 512 suppliers in the permit database could be found in AIMS. Similarly, only 136 out of 321 importers in the permit database could be found in AIMS. These results raise the possibility that either the importer and supplier names do not match well over the two databases, or a large number of licences are being sought for no reason. There are a number of reasons that this can happen.

- Some importers get a permit well in advance of imports ever happening or to cover gaps in supply that never eventuate, or choose not to import because of price fluctuations.
- Assuming supplier here is referring to "exporter" listed on the permit; the exporter might just be a shipping agent in the country of origin and might not match to the supplier in ICS/AIMS.
- Some permits will be for samples — and these will enter via air cargo — so will not be recorded not in AIMS.
- The supplier in AIMS may bear no relationship to the assessed manufacturer that is listed in the permit; i.e. for assessing the risk of feeds the manufacturer matters, and that information is not required for ICS/AIMS — DAFF would verify based on the documentation presented and not AIMS.

3.3.3 Alternatives

Subsequent review of this project by DAFF staff suggests that using a filter based on country and a few tariff codes/goods descriptions may have improved performance.

4

Data Mining

4.1 Introduction

The purpose of this chapter is to establish approaches and tools that are applicable for analyzing inspection data across pathways, that is, developing statistics and algorithms that might allow a collection of tariffs to be managed together. There are two motivations for this question. First, it is useful to identify those pathways that are likely to be suitable for a risk-based approach to pathway management. Such pathways will have a low underlying contamination rate, and ideally will be easy to divide into high-volume suppliers with good quarantine inspection histories and low-volume suppliers with a mix of good and bad histories, much as did the dried apricot pathway in Robinson et al. (2012). Second, suppliers are known to supply more than one type of commodity, and importers are known to import more than one type of commodity. It would be useful to determine whether or not it is valuable to impose an enhanced inspection regime on a supplier across all their pathways.

We note that when pathway failure rates are very low, it is very difficult to determine whether or not there is a pattern that can be exploited by CSP. Under these circumstances, it may be reasonable to use CSP even though the failures do not appear to have such a pattern. This would be because CSP "rewards" pathways that have a suitably clean quarantine inspection history by reducing intervention. This reward structure may provide motivation for suppliers or importers to keep their pathways clean. Further work is being undertaken in CEBRA Project 1305a.

The development of suitable tools and algorithms is still underway. For this report we present some useful summary statistics.

4.1.1 Data Preparation

Data were extracted from the AIMS database using the Tariff chapter 08 for a just over 5 year period, from 1 January 2007 to 21 March 2012 inclusive.

We defined inspection failures in two ways:

1. by cross-matching Incidents records that correspond to pests that are limited distribution, not in Australia, or uncertain (about 4100 fails at the quarantine entry level) — called a *quarantine fail.*

2. by finding all the records that contained at least one Inspection among the Direction Categories, and any of the following phrases within the Direction Results (using

case-insensitive matching): “not ok”, “failed”, or “breach” — called a *pathway fail*.

The overlap between these definitions can be shown as follows: this is a cross-tabulation of quarantine entries by inspection.

```
       Quarantine
Pathway  Fail  Pass
   Fail  3788  4323
   Pass   320 57883
```

This report uses the first definition.

4.1.2 Pathway Summary

Figure 4.1 sketches the inspection pathway for all consignments in tariff chapter 8, Fruit.

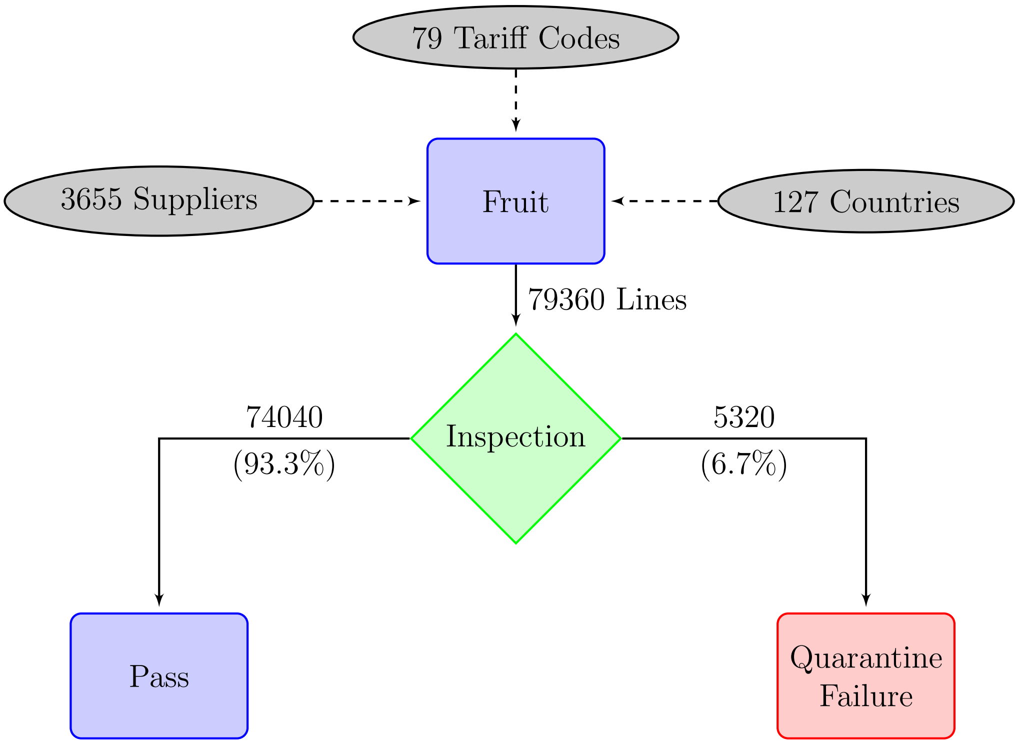

Figure 4.1: Fruit consignments flow chart with statistics for January 2007 to March 2012.

4.2 Analysis

This section provides a statistical overview of the data. The full dataset comprises 79360 lines with record creation dates ranging from January 2007 to March 2012, and comprises entries from 127 countries, 3655 suppliers, and 3150 importers. The entries are classified under 80 tariff codes and 8283 free-text goods descriptions.

Table 4.1 shows the pattern of imports by constituent and preparation. For example, only coconuts are desiccated, and all constituents are reported to have been imported at least once as frozen — including nuts (here, the tariff codes identified the product as being frozen (811 or 8119000), whereas the good description identified the consignment as nuts of some kind. Presumably the nut products with unknown preparation are fresh. Table 4.2 mimics the structure of Table 4.1 and shows the proportion of incoming lines that were inspected for each combination of constituent and preparation. We see for example that the great majority of nut products are inspected, whereas berries, avocados, pears and grapes are not inspected at so high a rate. Table 4.3 also mimics the structure of Table 4.1 and shows the proportion of inspected lines that failed for each combination of constituent and preparation. Failure rates are comparatively high for oranges, kiwi fruit, pomegranate, jack fruit and blueberries.

Table 4.4 describes the structure of the pathway from the point of view of exporting country. Among the major exporters, failure rates are particularly high for Fiji, Italy, New Zealand, and the USA. Tables 4.5 and 4.6 provide snapshots of the pathway from the point of view of the importers and the suppliers respectively. Tables 4.7 and 4.8 provide tariff-level information sorted decreasing by volume and increasing by contamination rate, respectively, and tables 4.9 and 4.10 provide the same information for the primary constituent as inferred from the goods description.

Table 4.1: Cross-tabulation of fruit lines by constituent and preparation, as reported in Goods Description or from tariff.

Constituent	Desicc	Dried	Fresh	Frozen	Peel	Powder	Shelled	Unknown	Total
(Rasp|Black)berries	0	72	16	439	0	4	0	0	531
Almonds	0	63	5	2	0	0	513	45	628
Apples	0	785	84	14	3	4	0	4	894
Apricots	0	1832	891	45	2	0	2	16	2788
Avocadoes	0	33	2848	24	0	0	0	959	3864
Bananas	0	319	3	459	0	26	0	475	1282
Blueberries	0	101	1916	190	0	0	0	210	2417
Brazil nuts	0	0	0	1	0	0	443	11	455
Cashew nuts	0	8	926	8	0	0	3947	90	4979
Cherries	0	105	3649	35	0	0	0	14	3803
Chestnuts	0	103	2	117	0	2	4	120	348
Coconuts	3094	192	358	792	0	82	0	1492	6010
Cranberries	0	155	7	10	0	2	0	27	201
Dates	0	1556	104	168	0	7	0	1940	3775
Durian	0	18	60	1031	0	0	0	1	1110
Figs	0	584	62	8	0	0	0	265	919
Goji berries	0	152	25	2	0	1	0	0	180
Grapes etc.	0	4007	8008	3	0	1	0	60	12079
Hazelnuts	0	0	25	1	0	0	710	18	754
Jackfruit	0	44	3	422	1	0	0	4	474
Kiwi fruit	0	15	4969	16	0	1	0	0	5001
Lemons & Limes	0	70	1254	5	51	15	0	203	1598
Longan	0	238	108	50	0	0	0	4	400
Macadamias	0	0	35	4	0	0	99	367	505
Mandarins etc.	0	21	518	2	30	0	0	122	693
Mangoes	0	504	867	130	1	21	0	575	2098
Misc. fruit	0	1506	88	712	61	15	0	419	2801
Oranges	0	90	4523	6	47	4	0	686	5356
Papaya	0	204	535	8	0	1	0	12	760
Passionfruit	0	0	22	168	0	2	0	0	192
Peaches	0	108	28	38	0	0	0	1	175
Pears	0	100	196	3	0	0	0	508	807
Persimmons	0	85	194	8	0	0	0	0	287
Pine nuts	0	5	15	1	0	0	0	514	535
Pineapples	0	200	27	28	0	5	0	148	408
Pistachio nuts	0	60	113	1	0	0	11	498	683
Plums etc.	0	1600	15	6	2	0	0	29	1652
Pomegranate	0	21	230	5	0	4	0	1	261
Strawberries	0	162	721	396	0	6	0	3	1288
Tamarind	0	493	15	66	2	1	0	6	583
Unknown	1	1333	415	791	212	60	2	923	3737
Walnuts	0	13	11	16	0	0	1845	164	2049
Total	3095	16957	33891	6231	412	264	7576	10934	79360

Table 4.2: Cross-tabulation of the percentage of fruit lines inspected by constituent and preparation, as reported in Goods Description or from tariff.

Constituent	Desicc	Dried	Fresh	Frozen	Peel	Powder	Shelled	Unknown	Overall
(Rasp|Black)berries		86.1	12.5	12.8		75.0			23.2
Almonds		98.4	80.0	0.0			95.9	82.2	94.7
Apples		89.2	71.4	14.3	100.0	25.0		100.0	86.1
Apricots		98.4	7.7	44.4	100.0		100.0	87.5	68.5
Avocadoes		9.1	7.7	0.0				7.3	7.5
Bananas		85.0	33.3	45.1		65.4		50.3	57.3
Blueberries		86.1	8.7	11.6				30.5	14.0
Brazil nuts				0.0			97.7	100.0	97.6
Cashew nuts		87.5	99.7	37.5			96.4	76.7	96.5
Cherries		86.7	43.1	37.1				21.4	44.1
Chestnuts		96.1	50.0	29.1		100.0	75.0	82.5	68.4
Coconuts	97.7	97.9	99.2	52.3		87.8		93.5	90.6
Cranberries		88.4	28.6	10.0		50.0		33.3	74.6
Dates		97.6	79.8	50.6		42.9		93.2	92.7
Durian		100.0	96.7	66.2				100.0	68.5
Figs		98.8	100.0	25.0				94.0	96.8
Goji berries		94.1	100.0	50.0		100.0			94.4
Grapes etc.		96.9	7.3	0.0		0.0		73.3	37.4
Hazelnuts			92.0	0.0			97.9	77.8	97.1
Jackfruit		75.0	100.0	66.8	0.0			75.0	67.7
Kiwi fruit		86.7	39.5	31.2		100.0			39.7
Lemons & Limes		94.3	62.5	20.0	62.7	100.0		62.1	64.1
Longan		91.2	98.1	50.0				100.0	88.0
Macadamias			94.3	75.0			100.0	95.1	95.8
Mandarins etc.		81.0	36.3	50.0	83.3			44.3	41.1
Mangoes		90.9	99.4	33.1	0.0	100.0		91.8	91.1
Misc. fruit		87.5	48.9	25.4	45.9	46.7		82.8	68.7
Oranges		80.0	70.1	0.0	80.9	100.0		52.2	68.0
Papaya		93.6	99.4	25.0		0.0		91.7	96.8
Passionfruit			100.0	41.7		50.0			48.4
Peaches		95.4	17.9	15.8				0.0	65.1
Pears		99.0	14.3	0.0				12.8	23.8
Persimmons		83.5	28.9	37.5					45.3
Pine nuts		80.0	93.3	0.0				99.2	98.7
Pineapples		90.5	100.0	35.7		60.0		87.2	85.8
Pistachio nuts		91.7	98.2	0.0			90.9	96.2	95.9
Plums etc.		90.4	80.0	50.0	50.0			58.6	89.5
Pomegranate		81.0	99.1	20.0		100.0		100.0	96.2
Strawberries		85.2	82.1	14.4		66.7		0.0	61.4
Tamarind		92.9	86.7	65.2	100.0	100.0		100.0	89.7
Unknown	100.0	88.5	84.1	33.8	77.8	41.7	100.0	76.4	72.1
Walnuts		100.0	90.9	50.0			98.9	72.0	96.3
Overall	97.7	93.2	39.4	41.0	71.8	70.5	97.2	72.6	63.7

Table 4.3: Percentage of fruit lines inspected that failed, tabulated by constituent and preparation.

Constituent	Desicc	Dried	Fresh	Frozen	Peel	Powder	Shelled	Unknown	Overall
(Rasp|Black)berries		0.0	0.0	0.0		0.0			0.0
Almonds		1.6	0.0				1.2	2.7	1.3
Apples		0.0	5.0	0.0	0.0	0.0		0.0	0.4
Apricots		0.7	0.0	0.0	0.0		0.0	0.0	0.6
Avocadoes		0.0	12.4					21.4	14.4
Bananas		1.1	0.0	4.8		0.0		0.8	2.0
Blueberries		0.0	53.6	0.0				35.9	33.0
Brazil nuts							0.7	9.1	0.9
Cashew nuts		0.0	0.7	0.0			1.4	1.4	1.2
Cherries		0.0	6.7	0.0				0.0	6.3
Chestnuts		0.0	0.0	52.9		0.0	0.0	4.0	9.2
Coconuts	0.2	51.1	40.6	1.9		1.4		31.6	12.8
Cranberries		0.0	50.0	0.0		0.0		22.2	2.0
Dates		2.0	6.0	0.0		0.0		0.4	1.2
Durian		0.0	6.9	0.1				0.0	0.7
Figs		2.1	88.7	0.0				20.5	13.3
Goji berries		0.0	0.0	0.0		0.0			0.0
Grapes etc.		0.6	14.8					9.1	2.5
Hazelnuts			0.0				0.7	14.3	1.0
Jackfruit		3.0	0.0	42.6				0.0	37.7
Kiwi fruit		0.0	52.2	0.0		0.0			51.7
Lemons & Limes		4.5	28.3	0.0	0.0	0.0		29.4	25.6
Longan		1.4	11.3	0.0				0.0	4.3
Macadamias			0.0	0.0			0.0	2.6	1.9
Mandarins etc.		0.0	22.3	0.0	8.0			20.4	19.3
Mangoes		5.7	13.1	2.3		0.0		12.1	10.7
Misc. fruit		2.4	44.2	1.7	0.0	0.0		0.3	2.8
Oranges		30.6	43.4		2.6	0.0		41.6	42.5
Papaya		1.0	30.6	0.0				0.0	22.4
Passionfruit			68.2	0.0		0.0			16.1
Peaches		1.0	0.0	0.0					0.9
Pears		0.0	7.1					4.6	2.6
Persimmons		0.0	58.9	0.0					25.4
Pine nuts		0.0	0.0					0.6	0.6
Pineapples		1.1	44.4	10.0		0.0		2.3	5.1
Pistachio nuts		1.8	0.0				0.0	0.2	0.3
Plums etc.		0.8	0.0	0.0	0.0			0.0	0.8
Pomegranate		0.0	60.5	0.0		0.0		100.0	55.4
Strawberries		0.0	11.0	0.0		0.0			8.2
Tamarind		1.5	23.1	0.0	0.0	0.0		0.0	1.9
Unknown	0.0	2.6	25.2	2.2	12.7	0.0	0.0	3.8	6.4
Walnuts		0.0	0.0	0.0			0.4	2.5	0.5
Overall	0.2	2.0	28.9	6.6	8.1	0.5	1.0	10.9	10.5

Table 4.4: Summary statistics by country. *Cons* is the number of lines, by which the table is ordered. *CF* is the number of lines with any number of contaminated fruit. *CR %* is the line-level quarantine failure rate amongst inspected lines. The *Importers* column reports the number of importers that have imported from each country during the time period. The *Suppliers* column reports the number of suppliers that have exported from each country during the time period. The *Fruits* column reports the number of fruit types that have exported from each country during the time period.

Country	Cons	Insp	CF	CR %	Importers	Suppliers	Fruits
United States	25083	12312	2297	18.7	505	517	36
New Zealand	12038	1989	583	29.3	361	211	36
China	6846	5477	114	2.1	416	638	39
Vietnam	5292	4760	34	0.7	167	319	20
Thailand	4924	4061	185	4.6	255	230	27
Turkey	4590	4507	37	0.8	140	222	22
Philippines	3253	2878	17	0.6	152	92	12
India	2120	1713	86	5.0	255	281	24
Iran Islamic Republic Of	1974	1951	11	0.6	107	119	15
Fiji	1527	1377	906	65.8	54	38	12
Italy	1158	1092	640	58.6	83	82	19
South Africa	731	660	1	0.2	61	58	17
Indonesia	680	668	4	0.6	57	60	8
Taiwan	582	509	8	1.6	80	75	17
Sri Lanka	532	430	10	2.3	85	69	12
Chile	487	314	2	0.6	64	81	19
Mexico	479	424	20	4.7	32	45	9
Netherlands	474	54	0	0.0	33	19	25
Australia	404	377	2	0.5	54	76	22
Saudi Arabia	387	377	2	0.5	343	83	7
United Arab Emirates	349	288	6	2.1	88	63	10
Israel	338	327	55	16.8	32	21	7
Hong Kong	330	301	0	0.0	74	53	23
Greece	328	309	2	0.6	35	43	10
Bolivia	327	319	1	0.3	25	37	9
Malaysia	327	261	4	1.5	58	54	15
Pakistan	308	303	20	6.6	55	53	15
Canada	259	82	0	0.0	42	38	16
France	240	195	60	30.8	42	31	18
Tonga	227	165	55	33.3	22	15	4
Brazil	222	169	4	2.4	46	52	15
Germany	209	168	1	0.6	43	23	21
Singapore	191	166	0	0.0	40	12	18
United Kingdom	188	108	1	0.9	62	29	23
Spain	170	154	24	15.6	30	25	15

Table 4.5: Summary statistics by importer. See Table 4.4 for explanations of the columns.

Importer	Cons	Insp	CF	CR %	Countries	Suppliers	Fruits
a	8719	4863	1002	20.6	16	30	22
b	7903	2593	946	36.5	18	114	26
c	3636	3603	29	0.8	20	205	25
d	2417	923	250	27.1	7	45	19
e	1751	1749	10	0.6	11	79	18
f	1456	1454	14	1.0	17	122	20
g	1443	1390	2	0.1	22	85	27
h	1323	396	67	16.9	8	34	18
i	1268	1254	6	0.5	18	50	18
j	1190	51	16	31.4	3	5	8
k	1129	516	137	26.6	9	68	21
l	919	49	2	4.1	1	7	3
m	896	66	10	15.2	3	22	6
n	834	826	2	0.2	4	11	9
o	829	814	1	0.1	20	44	21
p	799	352	148	42.0	5	28	9
q	741	701	0	0.0	17	55	23
r	692	118	0	0.0	28	27	22
s	651	311	103	33.1	7	36	11
t	621	249	109	43.8	4	18	4
u	615	600	62	10.3	7	13	13
v	601	79	25	31.6	2	13	4
w	590	529	38	7.2	5	19	9
x	566	532	5	0.9	12	101	16
y	514	492	0	0.0	16	52	21
z	497	271	101	37.3	9	37	22
A	497	57	10	17.5	2	9	6
B	495	477	425	89.1	1	2	9
C	472	296	0	0.0	17	33	17
D	471	142	77	54.2	7	22	13
E	452	395	1	0.3	8	16	15
F	448	103	41	39.8	1	9	2
G	431	408	3	0.7	12	44	20
H	412	138	59	42.8	8	34	22
I	405	382	0	0.0	14	20	14
J	386	144	53	36.8	6	35	15
K	370	369	340	92.1	1	1	10
L	367	354	2	0.6	10	36	13

Table 4.6: Summary statistics for fruit by supplier. See Table 4.4 for explanations of the columns.

Supplier	Cons	Insp	CF	CR %	Countries	Importers	Fruits
a	2944	2069	719	34.8	2	6	7
b	2308	41	1	2.4	1	2	3
c	2080	92	0	0.0	1	2	1
d	2066	1160	56	4.7	88	1476	36
e	1323	92	11	12.0	1	13	2
f	1301	652	251	38.5	2	1	6
g	1128	650	332	51.1	2	4	13
h	1116	551	140	25.4	2	2	10
i	898	32	13	40.6	1	1	7
j	878	53	3	5.7	1	22	3
k	874	873	6	0.7	4	3	3
l	874	192	97	50.5	2	20	14
m	849	339	81	23.9	3	3	12
n	782	288	21	7.3	1	6	1
o	616	56	12	21.4	2	10	5
p	601	591	8	1.4	3	2	4
q	596	594	0	0.0	1	1	1
r	524	523	0	0.0	1	15	8
s	493	41	2	4.9	1	12	1
t	464	462	0	0.0	2	3	1
u	423	423	0	0.0	1	5	1
v	404	379	340	89.7	1	4	9
w	395	395	3	0.8	1	1	1
x	392	249	54	21.7	2	2	6
y	382	377	0	0.0	3	5	7
z	379	315	165	52.4	1	4	1
A	377	36	6	16.7	1	5	1
B	371	370	340	91.9	1	2	10
C	370	133	45	33.8	2	3	8
D	363	10	1	10.0	1	5	2
E	361	14	0	0.0	10	1	17
F	358	349	2	0.6	5	1	4
G	348	13	2	15.4	1	3	2
H	341	341	220	64.5	1	1	1
I	322	31	4	12.9	2	13	5
J	304	176	0	0.0	1	2	9
K	304	110	15	13.6	1	10	6
L	303	299	1	0.3	3	8	11

Table 4.7: Summary statistics for fruit by tariff, ordered from high to low by volume. Minimum 400 lines. See Table 4.4 for explanations of the columns.

Tariff	Cons	Insp	CF	CR %	Countries	Suppliers	Importers
8061000	7602	350	31	8.9	2	71	34
8051000	5290	3600	1552	43.1	18	100	86
8105000	4906	1941	1012	52.1	8	113	84
8013200	4879	4733	59	1.2	20	350	112
8119000	4651	2180	165	7.6	55	426	271
8134000	4609	4198	64	1.5	64	788	506
8062000	3848	3801	25	0.7	29	357	209
8044000	3837	290	42	14.5	3	62	51
8092000	3657	1588	105	6.6	7	70	51
8041000	3303	3106	42	1.4	34	505	358
8011100	3162	3078	8	0.3	28	186	144
8011900	2111	1936	676	34.9	23	184	140
8045000	1898	1802	201	11.2	24	254	223
8023200	1854	1840	7	0.4	22	118	76
8131000	1797	1770	11	0.6	27	214	149
8109000	1607	786	365	46.4	20	133	98
8104000	1507	180	64	35.6	9	45	35
8055000	1501	955	260	27.2	23	117	97
8029000	1311	1246	17	1.4	36	330	236
8132000	1163	1067	9	0.8	25	131	109
8091000	899	81	2	2.5	9	56	57
8042000	898	879	118	13.4	24	177	137
8135000	856	735	5	0.7	32	147	129
8133000	755	684	0	0.0	17	99	65
8082000	748	112	8	7.1	7	69	45
8022200	732	715	5	0.7	13	64	48
8101000	727	594	65	10.9	6	30	25
8112000	671	96	0	0.0	23	103	59
8052000	657	253	53	20.9	8	36	27
8025000	640	622	2	0.3	15	70	58
8021200	572	550	7	1.3	26	142	95
8072000	511	505	136	26.9	4	19	19
8039000	476	404	4	1.0	24	94	77
8054000	448	262	59	22.5	4	32	20
8012200	442	432	3	0.7	12	63	34
8026000	427	416	9	2.2	14	46	30
803	407	131	1	0.8	44	29	298

Table 4.8: Summary statistics for fruit by tariff, ordered from low to high by contaminatio nn rate. Minimum 250 inspected lines. See Table 4.4 for explanations of the columns.

Tariff	Cons	Insp	CF	CR %	Countries	Suppliers	Importers
8133000	755	684	0	0.0	17	99	65
8011100	3162	3078	8	0.3	28	186	144
804	382	327	1	0.3	32	4	368
8025000	640	622	2	0.3	15	70	58
8023200	1854	1840	7	0.4	22	118	76
8131000	1797	1770	11	0.6	27	214	149
8062000	3848	3801	25	0.7	29	357	209
8135000	856	735	5	0.7	32	147	129
8012200	442	432	3	0.7	12	63	34
8022200	732	715	5	0.7	13	64	48
8132000	1163	1067	9	0.8	25	131	109
8039000	476	404	4	1.0	24	94	77
8013200	4879	4733	59	1.2	20	350	112
8021200	572	550	7	1.3	26	142	95
8041000	3303	3106	42	1.4	34	505	358
8029000	1311	1246	17	1.4	36	330	236
8134000	4609	4198	64	1.5	64	788	506
8026000	427	416	9	2.2	14	46	30
8043000	364	330	17	5.2	18	71	70
8092000	3657	1588	105	6.6	7	70	51
8119000	4651	2180	165	7.6	55	426	271
8140000	374	294	23	7.8	20	98	64
8061000	7602	350	31	8.9	2	71	34
8101000	727	594	65	10.9	6	30	25
8045000	1898	1802	201	11.2	24	254	223
8042000	898	879	118	13.4	24	177	137
8044000	3837	290	42	14.5	3	62	51
8052000	657	253	53	20.9	8	36	27
8054000	448	262	59	22.5	4	32	20
8072000	511	505	136	26.9	4	19	19
8055000	1501	955	260	27.2	23	117	97
8011900	2111	1936	676	34.9	23	184	140
8051000	5290	3600	1552	43.1	18	100	86
8109000	1607	786	365	46.4	20	133	98
8105000	4906	1941	1012	52.1	8	113	84

Table 4.9: Summary statistics for fruit by constituent, ordered from high to low by volume. See Table 4.4 for explanations of the columns.

Constituent	Cons	Insp	CF	CR %	Countries	Suppliers	Importers
Grapes etc.	12079	4513	115	2.5	39	478	305
Coconuts	6010	5447	697	12.8	37	372	312
Oranges	5356	3644	1550	42.5	27	132	121
Kiwi fruit	5001	1984	1026	51.7	12	131	101
Cashew nuts	4979	4806	61	1.2	23	364	165
Avocadoes	3864	291	42	14.4	7	68	59
Cherries	3803	1678	106	6.3	18	108	85
Dates	3775	3498	43	1.2	44	515	721
Unknown	3737	2694	173	6.4	93	797	982
Misc. fruit	2801	1923	54	2.8	69	559	613
Apricots	2788	1910	12	0.6	32	288	233
Blueberries	2417	339	112	33.0	14	110	79
Mangoes	2098	1912	204	10.7	31	310	291
Walnuts	2049	1973	10	0.5	25	133	130
Plums etc.	1652	1479	12	0.8	36	213	171
Lemons & Limes	1598	1024	262	25.6	34	147	138
Strawberries	1288	791	65	8.2	29	157	112
Bananas	1282	735	15	2.0	46	150	362
Durian	1110	760	5	0.7	4	64	54
Figs	919	890	118	13.3	25	186	152
Apples	894	770	3	0.4	23	139	131
Pears	807	192	5	2.6	14	98	62
Papaya	760	736	165	22.4	16	62	64
Hazelnuts	754	732	7	1.0	13	72	56
Mandarins etc.	693	285	55	19.3	10	53	41
Pistachio nuts	683	655	2	0.3	16	77	66
Almonds	628	595	8	1.3	32	151	117
Tamarind	583	523	10	1.9	10	124	100
Pine nuts	535	528	3	0.6	11	117	82
(Rasp|Black)berries	531	123	0	0.0	22	104	63
Macadamias	505	484	9	1.9	17	48	46
Jackfruit	474	321	121	37.7	8	64	57
Brazil nuts	455	444	4	0.9	12	67	37
Pineapples	408	350	18	5.1	23	88	88
Longan	400	352	15	4.3	5	75	69

Table 4.10: Summary statistics for fruit by constituent, ordered from low to high by compliance rate. See Table 4.4 for explanations of the columns.

Constituent	Cons	Insp	CF	CR %	Countries	Suppliers	Importers
(Rasp|Black)berries	531	123	0	0.0	22	104	63
Pistachio nuts	683	655	2	0.3	16	77	66
Apples	894	770	3	0.4	23	139	131
Walnuts	2049	1973	10	0.5	25	133	130
Pine nuts	535	528	3	0.6	11	117	82
Apricots	2788	1910	12	0.6	32	288	233
Durian	1110	760	5	0.7	4	64	54
Plums etc.	1652	1479	12	0.8	36	213	171
Brazil nuts	455	444	4	0.9	12	67	37
Hazelnuts	754	732	7	1.0	13	72	56
Dates	3775	3498	43	1.2	44	515	721
Cashew nuts	4979	4806	61	1.2	23	364	165
Almonds	628	595	8	1.3	32	151	117
Macadamias	505	484	9	1.9	17	48	46
Tamarind	583	523	10	1.9	10	124	100
Bananas	1282	735	15	2.0	46	150	362
Grapes etc.	12079	4513	115	2.5	39	478	305
Pears	807	192	5	2.6	14	98	62
Misc. fruit	2801	1923	54	2.8	69	559	613
Longan	400	352	15	4.3	5	75	69
Pineapples	408	350	18	5.1	23	88	88
Cherries	3803	1678	106	6.3	18	108	85
Unknown	3737	2694	173	6.4	93	797	982
Strawberries	1288	791	65	8.2	29	157	112
Mangoes	2098	1912	204	10.7	31	310	291
Coconuts	6010	5447	697	12.8	37	372	312
Figs	919	890	118	13.3	25	186	152
Avocadoes	3864	291	42	14.4	7	68	59
Mandarins etc.	693	285	55	19.3	10	53	41
Papaya	760	736	165	22.4	16	62	64
Lemons & Limes	1598	1024	262	25.6	34	147	138
Blueberries	2417	339	112	33.0	14	110	79
Jackfruit	474	321	121	37.7	8	64	57
Oranges	5356	3644	1550	42.5	27	132	121
Kiwi fruit	5001	1984	1026	51.7	12	131	101

4.3 Discussion

Table 4.8 identifies several tariff codes that have very high compliance rates and several that have very poor compliance rates. Of note are those codes that have *high* inspection rates and *low* interception rates, and also those tariffs that have *low* inspection rates and *high* interception rates.

In order to impose some structure on the data, we chose the following approach to classifying the pathways. If the pathway contamination rate were less than 4% and the inspection rate greater than 70%, then the pathway is *low residual risk*, or *low risk*, and is eligible for further analysis, e.g. using the simulation approach presented in Robinson et al. (2012). If the pathway contamination rate were more than 10% and the inspection rate less than 90% then the pathway is defined as being *high residual risk*, or *high risk*. Here, the risk refers to the risk presented by the pathway based on the contamination rate of the pathway and the current management approach. Therefore, the risk is defined by the compliance rate, not the biological risk presented by the types of contamination that the pathway may be suspected to carry. The cutoff levels are arbitrary, and a pathway with high compliance may present a higher risk to biosecurity than a pathway with lower compliance simply based on the types of contamination in the pathways. Figure 4.2 provides a graphical summary of the tables.

We note that use of the goods description for data mining has disadvantages, for example, it is subject to inconsistent spelling, vague specification, etc. Some types of descriptions could be considered at the same level as tariffs, that is, they are both open to potential abuse. A precautionary principal can be applied, so for example if dried dates are safer than fresh dates then we may assume that the dates are fresh unless the goods description includes the word `DRIED`.

4.4 Dates, *Redux*

The contamination results for dates, dried and fresh, in this chapter differ from the results in the previous chapter by approximately three-fold. That is, when analyzed alone the dates pathway was quite clean, with a recent failure rate of less than 0.5%, whereas when analyzed as part of the Chapter 08 data (fruit tariffs), the failure rate was close to 1.5%.

We did some drilling into the data to determine the reason for this discrepancy. The likely cause is that incidents data have been recorded against the quarantine entry, rather than the line within the quarantine entry. When we merged the AIMS and Incidents data, we were able to merge only to the entry level. Numerous entries comprise more than one line, and often, the incident is raised against a line other than that under analysis. For example, the following table counts the commodity descriptions from Incidents that correspond to the quarantine entries that were classified in the fruit analysis as being dates, fresh or otherwise, from July 2008 onwards (corresponding to the analysis period used in Section 3.1.2).

```
desc
                                           ANIMAL RESIDUE
                         4                              1
           BREAD ALL TYPES                      CARDBOARD
                         1                              4
         CEREALS PROCESSED   FISH / FISH PRODUCTS DRIED
```

Table 4.11: Summary statistics for frequently inspected (at least 250 inspections in whole dataset) fruit by tariff, ordered from high to low by compliance rate, with descriptions. See Table 4.4 for explanations of the columns.

Tariff	Cons	Insp	CR %	Preparation	Content
8133000	755	684	0.0	Dried	Apples
8011100	3162	3078	0.3	Desiccated	Coconuts
804	382	327	0.3		Dates Etc
8025000	640	622	0.3		Pistachios
8023200	1854	1840	0.4	Shelled	Walnuts
8131000	1797	1770	0.6	Dried	Apricots
8062000	3848	3801	0.7	Dried	Grapes
8135000	856	735	0.7		Fruit Mixtures
8012200	442	432	0.7	Shelled	Brazil Nuts
8022200	732	715	0.7	Shelled	Hazelnuts
8132000	1163	1067	0.8	Dried	Prunes
8039000	476	404	1.0		Bananas
8013200	4879	4733	1.2	Shelled	Cashews
8021200	572	550	1.3	Shelled	Almonds
8041000	3303	3106	1.4		Dates
8029000	1311	1246	1.4		Other Nuts
8134000	4609	4198	1.5	Dried	Other Fruit
8026000	427	416	2.2		Macadamias
8043000	364	330	5.2		Pineapples
8092000	3657	1588	6.6	Fresh	Cherries
8133000	4651	2180	7.6	Frozen	Other Fruit Etc.
8011100	374	294	7.8	Peel	Citrus Or Melon
804	7602	350	8.9	Fresh	Grapes
8025000	727	594	10.9	Fresh	Strawberries
8023200	1898	1802	11.2		Guavas Etc.
8131000	898	879	13.4		Figs
8062000	3837	290	14.5		Avocados
8135000	657	253	20.9		Mandarins
8012200	448	262	22.5		Grapefruit
8022200	511	505	26.9	Fresh	Pawpaws
8132000	1501	955	27.2		Lemons
8039000	2111	1936	34.9		Other Coconuts Etc
8013200	5290	3600	43.1		Oranges
8021200	1607	786	46.4	Fresh	Other Fruit
8041000	4906	1941	52.1	Fresh	Kiwi

Table 4.12: Summary statistics for frequently inspected (at least 250 inspections in whole dataset) fruit by tariff, ordered from high to low by compliance rate, with descriptions, for inspections since January 2011. See Table 4.4 for explanations of the columns. A heuristic risk rating is added; see text for explanation.

Tariff	Cons	Insp	CR %	Preparation	Content	Risk
8012200	89	87	0.0	Shelled	Brazil Nuts	Low
8022200	198	186	0.0	Shelled	Hazelnuts	Low
8026000	115	113	0.0		Macadamias	Low
8061000	2107	5	0.0	Fresh	Grapes	Moderate
8132000	330	303	0.0	Dried	Prunes	Low
8133000	194	182	0.0	Dried	Apples	Low
8135000	188	146	0.0		Fruit Mixtures	Low
8023200	472	466	0.2	Shelled	Walnuts	Low
8131000	440	423	0.2	Dried	Apricots	Low
8011100	840	806	0.2	Desiccated	Coconuts	Low
8062000	1025	1005	0.3	Dried	Grapes	Low
8029000	370	340	0.6		Other Nuts	Low
8021200	172	158	0.6	Shelled	Almonds	Low
8025000	149	143	0.7		Pistachios	Low
8013200	1043	1010	1.2	Shelled	Cashews	Low
8041000	848	801	1.2		Dates	Low
804	88	80	1.2		Dates Etc	Low
8134000	1170	1021	1.9	Dried	Other Fruit	Low
8042000	214	205	2.9		Figs	Low
8039000	133	102	3.9		Bananas	Low
8012200	86	76	3.9		Pineapples	Low
8022200	1444	629	4.9	Frozen	Other Fruit Etc.	Moderate
8026000	428	394	10.9		Guavas Etc.	Moderate
8061000	971	146	11.0	Fresh	Cherries	High
8132000	108	87	13.8	Peel	Citrus Or Melon	High
8133000	1163	31	19.4		Avocados	High
8135000	164	78	21.8		Grapefruit	High
8023200	285	116	23.3		Mandarins	High
8131000	193	192	26.0	Fresh	Pawpaws	Moderate
8011100	630	571	27.0		Other Coconuts Etc	Moderate
8062000	350	201	30.3		Lemons	High
8029000	1800	1250	38.3		Oranges	High
8021200	1207	533	44.7	Fresh	Kiwi	High
8025000	80	2	50.0	Fresh	Strawberries	High
8013200	546	201	51.7	Fresh	Other Fruit	High

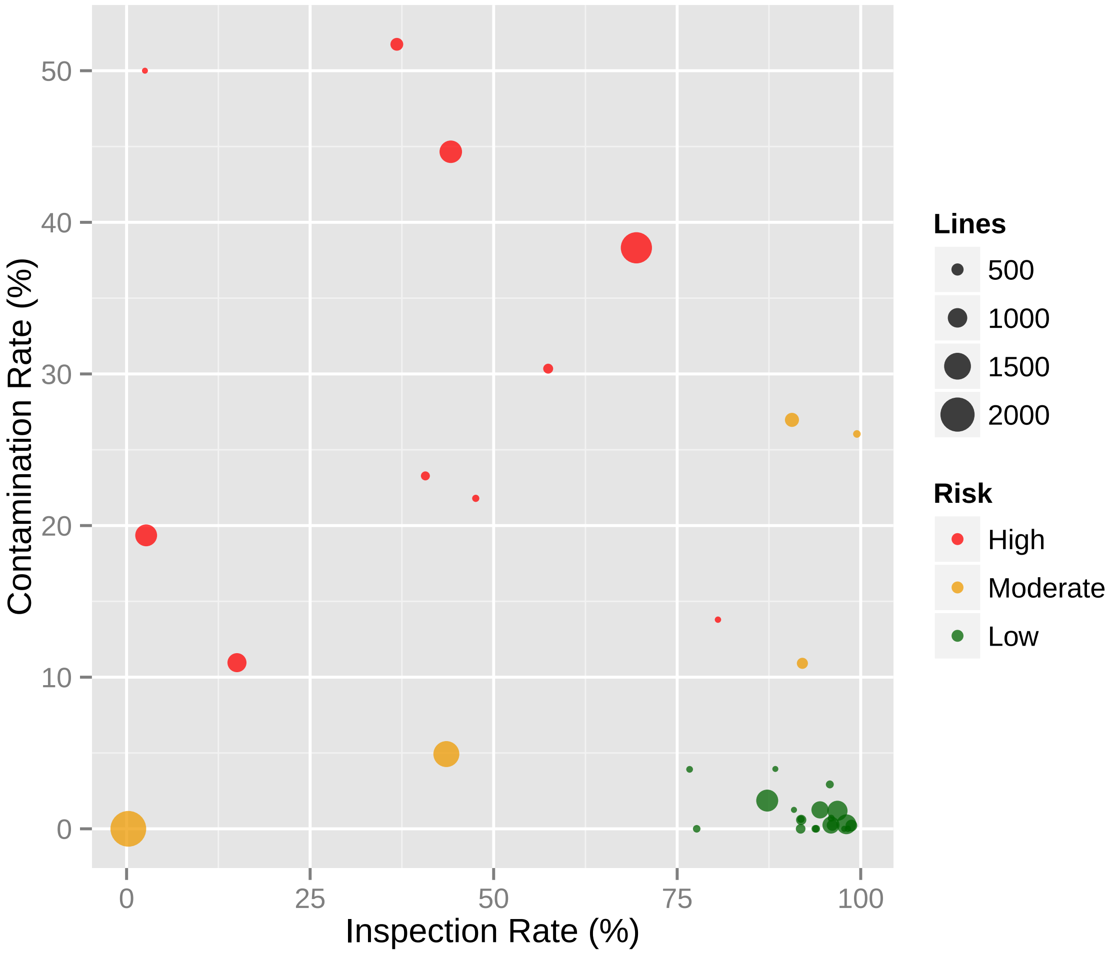

Figure 4.2: Graphical summary of tariff-specific analysis of inspection rates and contamination rates. See text for description of classification.

```
                        1                             1
        FRUIT DATES DRIED              FRUIT DATES FRESH
                        6                             3
  FRUIT POMEGRANATE FRESH    GRAIN / SEED BARLEY DRIED
                        1                             2
GRAIN / SEED RICE POLISHED    HERB / SPICE GINGER DRIED
                        3                             1
   HERB / SPICE MIXED DRIED        HERBAL MEDICINE DRIED
                        1                             3
       Pot Pourri - Dried                          SEED
                        1                             1
 Seeds For Sowing - Ocimum                 TIMBER PALLET
                        1                             1
 VEGETABLE ASPARAGUS FRESH VEGETABLE BAMBOO SHOOT DRIED
                        1                             1
  VEGETABLE MUSHROOM DRIED
                        2
```

The tabulation clearly shows that a small proportion of the lines (22%) are described as dates, so the analysis is very conservative.

The sole remedy is manually cross-checking the commodity description against the goods descriptions of each of the lines. This is time consuming. In future, it would be preferred if the line number corresponding to the interception were recorded on Incidents.

5

Residual Elements

5.1 Determination of Training Needs

Plant Biosecurity is negotiating with ABARES to provide ongoing analytical support. ACERA has been working with ABARES staff to provide code and suitable documentation. The anticipated training of DAFF staff to use the algorithms and code that arise from this project now seems unnecessary. Hence, this deliverable was eliminated after discussion with the project leaders and sponsor.

5.2 Identification of Processes Needed to Measure Biological Risk

The distinction between actionable and non-actionable contamination is still blurry. This deliverable of the project requires the innovations being developed by ACERA Project 1101E, *Sampling for Invasives*, which is underway.

Bibliography

Dodge, H. (1943). A sampling inspection plan for continuous production. *The Annals of Mathematical Statistics*, 14(3):264–279.

Dodge, H. F. and Torrey, M. N. (1951). Additional continuous sampling plans. *Industrial Quality Control*, 7(5):7–12.

Gelman, A., Jakulin, A., Pittau, M. G., and Su, Y.-S. (2008). A weakly informative default prior distribution for logistic and other regression models. *The Annals of Applied Statistics*, 2(4):1360–1383.

Gelman, A., Su, Y.-S., Yajima, M., Hill, J., Pittau, M. G., Kerman, J., and Zheng, T. (2011). *arm: Data Analysis Using Regression and Multilevel/Hierarchical Models*. R package version 1.4-14.

Hilbe, J. M. (2009). *Logistic Regression Models*. Chapman and Hall/CRC, Boca Raton.

McCullagh, P. and Nelder, J. (1989). *Generalized Linear Models*. Chapman and Hall, London, 2nd edition.

Robinson, A., Bell, J., Woolcott, B., and Perotti, E. (2012). DAFF Biosecurity: Plant-Product Pathways. Technical Report 1001J 1, Australian Centre of Excellence for Risk Analysis.

Robinson, A., Cannon, R., and Mudford, R. (2011). AQIS Quarantine Operations Risk Return: Performance Indicators. Technical Report 1001I 1, Australian Centre of Excellence for Risk Analysis.

Warnes, G. R. (2011a). *gmodels: Various R programming tools for model fitting*. R package version 2.15.1.

Warnes, G. R. (2011b). *gplots: Various R programming tools for plotting data*. R package version 2.10.1.

Appendix A

Post-Intervention Compliance (PIC)

Here we provide a brief summary of how to compute the PIC performance indicator as used in this report. For more details and other variations, see Robinson et al. (2011). This text is identical to that provided in Robinson et al. (2012).

The calculation is based on division of the items in the pathway into two strata: *enhanced inspection* and *monitored*. We estimate the number of missed consignments for each stratum and then sum them to obtain the overall count, which is used to compute the PIC.

We define, for each stratum,

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

i as the number of units actually inspected,

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

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

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

$\hat{l}$ as the estimated leakage count of units *that were inspected*,

$\hat{L}$ as the estimated pathway-level leakage count,

$\hat{a}$ as the approach count, and

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

The post-intervention compliance (PIC) is calculated as

$$\text{PIC} = \frac{v - \hat{L}}{v} \tag{A.1}$$

If a leakage survey had been done then we would estimate $\hat{l} = i \times y/n$ and then estimate $\hat{e} = b/(b + \hat{l})$. In the absence of a leakage survey, we have assumed that inspection effectiveness e is known, and the same for each stratum. Given e, the inspection-level leakage $\hat{l}$ is

$$\hat{l} = \frac{b}{e} - b = b \times \left(\frac{1}{e} - 1\right) \tag{A.2}$$

Then, the estimated approach count for each stratum (monitored: $k = 1$, enhanced inspection: $k = 2$) is

$$\hat{a}_k = \frac{b_k}{e} \times \frac{v_k}{i_k} \tag{A.3}$$

where $v_2 = i_2$ for the enhanced inspection stratum, and $i_1 < v_1$ for the sampled stratum. Then

$$\hat{L}_k = \hat{a}_k - b_k = b_k \times \left(\frac{v_k}{i_k \times e} - 1 \right) \tag{A.4}$$

and

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

The final step is then to compute the quantity in equation (A.1).

Note that for these simulation experiments, we assume that the effectiveness is 0.9, so $\hat{l} = 0.11 \times b$, and

$$\hat{L}_k = b_k \times \left(\frac{v_k}{i_k \times 0.9} - 1 \right) \tag{A.6}$$