Evaluating the Health of Australia's Biosecurity System

Summary of Technical Report: Key Performance Indicators of "Anticipate" and "Prevent" Activities

Stephen E. Lane[1], John B. Baumgartner[1], and Andrew P. Robinson[1]

[1]CEBRA, The University of Melbourne

June 16, 2020

cebra

Centre of Excellence for Biosecurity Risk Analysis

Contents

List of Figures

1. Introduction

The Department of Agriculture, Water, and the Environment (the department) and CEBRA have collaborated on a 3-year project to develop a framework and candidate performance indicators for measuring the health of the biosecurity system. This technical report supports the final report of CEBRA Project 170714, *Health of the Biosecurity System*, and should be read accompanying that report.

The department has long recognised the importance of risk-based approaches to managing the biosecurity risk of various pathways within which biosecurity risk material may enter Australia. Importantly, pathway biosecurity risk management mostly involves the imposition of regulatory requirements. The purpose of border inspection is primarily to verify whether the pathway is compliant with regulations.

Schneider and Arndt (2019) described the biosecurity system as comprising actions classified under seven categories, namely anticipate, prevent, screen, prepare, detect, respond, and recover & adapt. Here we report the use of proposed performance indicators that are computed using departmental interception data to assess some aspects of the performance of the biosecurity system to *anticipate* and *prevent* the arrival of biosecurity risk material at our borders[1]. We focus on an international pathway for which the needed data are readily available, unlike most other pathways.

Performance indicators that can be used to help manage pathway risk have been developed by a number of CEBRA reports (Robinson et al., 2011; Robinson et al., 2013; Hoffmann et al., 2016), with four particular indicators recommended as follows:

- Approaching Compliance (AC), the proportion of items arriving at the border that is compliant;
- Residual Compliance (RC), the proportion of items that is compliant after intervention;
- Non-compliance Effectiveness (NCE), the proportion of non-compliant items that are identified and corrected or removed; and
- Hit Rate, the proportion of interventions that identify non-compliance.

In this report we demonstrate the use of two of these performance indicators (namely approaching compliance and non-compliance effectiveness) which we describe in greater detail in sections 2.1 and 2.2 respectively. We focus mostly on approaching compliance in order to reflect the performance of the biosecurity system with respect to its capacity to *Anticipate* and *Prevent* the arrival of biosecurity risk material at our borders.

Robinson et al. (2011), Robinson et al. (2013), and Hoffmann et al. (2016) briefly discuss how confidence intervals may be used by an analyst to suggest when the

[1]The focus of this report is unusual in that border inspection data are more commonly used to assess performance in *Screening*.

pathway may require some form of intervention, but none of these reports develop confidence intervals for comparisons over time, nor for aggregation at different levels of the pathway.

In order to estimate the performance indicators, we require either a random sample of inspection results on the pathway, or full inspection of units on the pathway. Moreover, we must either assume that inspection efficacy is 100% (that is, if a unit is contaminated, it will always be found), or have a sufficient estimate of the efficacy of inspection. When a pathway does not have every unit inspected, and those that are selected have not been sampled completely at random (e.g., when units have been selected based on profiling[2] and risk assessment), an endpoint survey may be used to enable estimation of the performance indicators.

An endpoint survey is an inspection of units that have *already been cleared* at the border. Endpoint surveys allow the estimation of the leakage rate, which is the rate of biosecurity risk material that is not intercepted upon initial inspection at the border. Thus, endpoint surveys are required for two important reasons:

1. The volume of some pathways is so large that not every unit is inspected. Hence, *units* on the pathway are *profiled* for inspection, leading to a biased sample for estimation of the performance indicators; and

2. Inspection of units is not always perfect, which leads to *leakage* (i.e. some non-compliant units not being found during inspection).

We use multiple performance indicators to monitor the health of various pathways, as each indicator is appropriate for various parts of the pathway. As an example, leakage rates and non-compliance effectiveness provide information on border activities, whilst approaching compliance rates provide information on pre-border and offshore biosecurity activities.

Each indicator can be further classified according to the level of the system that it is monitoring. For example, all air passengers is a pathway. However, this pathway can be broken down by location, where passengers entering a location is also a pathway. Each of these pathways may have different risks of biosecurity risk material, so that estimating performance indicators for each is necessary. And second, within each pathway, there may be many different intervention methods. Each of the intervention methods (for example, manual inspection and canine inspection) are likely to have different levels of efficacy, meaning that leakage rates will differ between them.

Monitoring what may account to several hundred indicators—depending on the different combinations of pathway and intervention methods—is not possible for a high-level report on the health of the pathway. Accordingly, we require a framework that accounts for the various combinations of pathways and intervention methods, that will simultaneously produce a smaller number of key headline indicators to monitor.

This technical report proposes a framework that allows the monitoring of multiple performance indicators at a higher level, whilst also being flexible enough to monitor all of the component pathways. Section 2 describes the performance indicators in more detail, including how they can be calculated. Section 3 describes the estimation and modelling of leakage rates, and proposes a general framework for monitoring performance indicators. Case studies of two pathways were undertaken and described

[2]Here, profiling is the application of risk-based intervention by focusing extra inspection resources on pathways that have a history of higher biosecurity risk.

in detail in a full version of this report provided to the department. While the details of these case studies are omitted from this summary report due to data sensitivities, the underlying models are described in Section A. Section 4 discusses some of the issues related to measurement and monitoring of the indicators.

2. Calculation of Performance Indicators

In this section we describe the performance indicators in more detail, including descriptions of how they may be calculated. This section is based on Hoffmann et al. (2016), to which we refer readers for more detail. Figure 2.1 provides a simplified description of how units interact with the department, and is derived from similar figures in Hoffmann et al. (2016).

The flow of units through the pathway is as follows: V_I units are required to be inspected (e.g. due to profiling, left side of Figure 2.1), and V_{NI} units are not required to be inspected (right side of Figure 2.1). As it is expected that not all non-compliance will be detected by inspection (and there will be some units that aren't inspected, yet are actually non-compliant), an endpoint survey is used to estimate how much leakage (see Section 3) has occurred. The endpoint survey is conducted on a random sample of N_1 inspected units that were deemed to be compliant, and N_2 units that were not inspected. In the inspected units, the endpoint inspection finds that NC_1 were actually non-compliant, whilst in the units that were not inspected, the endpoint survey finds that NC_2 units were actually non-compliant.

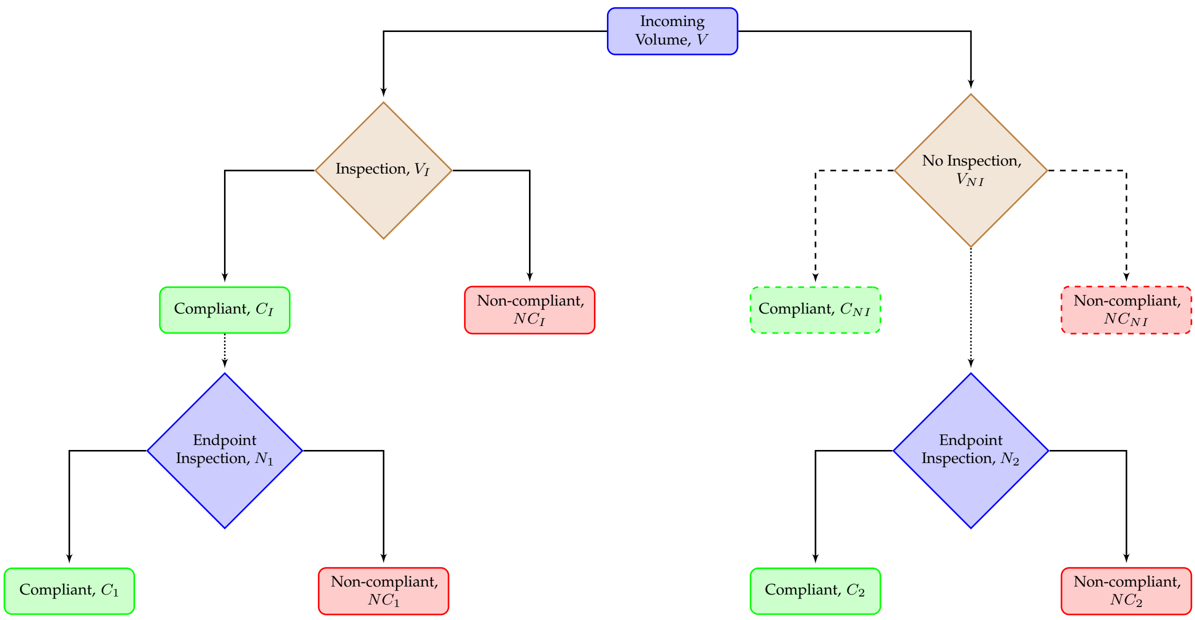

Figure 2.1.: Flowchart depicting how units interact with the department (simplified). Dotted lines denote that endpoint inspections occur only on a (random) sample of units. Volumes at each stage are provided after the stage label (e.g., the incoming volume is V). The true number of compliant and non-compliant units that were not inspected (C_{NI} and NC_{NI} respectively) are unknown.

2.1. Approaching Compliance Rate

The approaching compliance rate is the proportion of units that are compliant before any intervention by the department, for example, inspection. Figure 2.1 shows the number of units that are compliant prior to interaction with the department is some fraction of C_I, which is the number of units that are labelled as compliant following inspection, plus C_{NI} (which is unknown). C_{NI} can be estimated by multiplying the number of units that weren't inspected (V_{NI}), by the fraction of units that were truly compliant following the endpoint inspection (C_2/N_2). Thus, we can estimate the total number of compliant units, say $\widehat{C}$, prior to interaction with the department as:

$$\widehat{C} = C_I \frac{C_1}{N_1} + V_{NI} \frac{C_2}{N_2} \tag{2.1}$$

We can then write down an estimate of the approaching compliance rate AC as

$$\widehat{AC} = \frac{\widehat{C}}{V} \tag{2.2}$$

2.2. Non-compliance Effectiveness

The non-compliance effectiveness (or simply effectiveness) is the number of units that were detected as non-compliant following inspection, divided by the total number of non-compliant units. Effectiveness thus provides a measure of how well our profiling and inspection is working: if effectiveness were 1, then there would be zero leakage, i.e. our profiling and inspection were perfect. The total number of non-compliant units is made up of those units deemed non-compliant following inspection (NC_I); the units that we 'missed' following inspection, i.e. the number of units that should have been deemed non-compliant following inspection; and the number of units that we didn't inspect, but were actually non-compliant. Putting these together, we can write down an estimate of the total number of non-compliant units, say $\widehat{NC}$, as

$$\widehat{NC} = NC_I + C_I \frac{NC_1}{N_1} + V_{NI} \frac{NC_2}{N_2} \tag{2.3}$$

We can then write down an estimate of the effectiveness, NCE as

$$\widehat{NCE} = \frac{NC_I}{\widehat{NC}} \tag{2.4}$$

2.3. A Note on Leakage

Sections 2.1 and 2.2 demonstrated how to calculate the performance indicators that we use in this report. Endpoint surveys provided important information in each of these calculations, due to the (likely) presence of leakage; that is, inspection is unlikely to be perfect, so we miss some non-compliant units, and profiling is unlikely to be perfect, so we don't inspect some units that are actually non-compliant.

The estimation of leakage rates is thus pivotal to the estimation of the performance indicators. Estimation of the leakage rates will be developed in the following sections.

3. Pathway Leakage Rates

A key component of estimating the performance indicators (Section 2) is the estimation of the leakage rate. In this chapter we demonstrate estimation of *point estimates* of the indicators (Section 3.1). We then develop a framework for monitoring the performance indicators over time (Section 3.2) and then provide some arguments for why statistical modelling is required for this framework. Finally we demonstrate how we can monitor the indicators at various levels of the pathway via pathway volume aggregation (Section 3.4).

3.1. Estimating Leakage Rates

Leakage rates may vary significantly over time, as well as within different intervention method pathways (also called channels). This can be for a range of reasons, including the types of risk material presented, and the change in inspection effort. Pathways can be disaggregated by the location the items pass through (the *entry point*), the intervention method used, and the item classes. For convenience, we refer to a combination of entry point, item class (if relevant), and intervention method as a *sub-pathway*. The leakage rate is calculated as the fraction of non-compliant units following the endpoint inspection; from Figure 2.1, the estimated leakage rate following inspection would be calculated as NC_1/N_1.

Leakage rates can vary across sub-pathways, as well as within sub-pathways over time. If only point estimates of the leakage rates were to be used, this variability may lead to false outcomes when monitoring the health of the sub-pathway. Furthermore, there is no way to handle missing data when using point estimates. Thus, a framework for monitoring performance indicators derived from leakage rates will clearly need to account for both missing data, and the variability not only between sub-pathways, but also within sub-pathways over time.

In some cases the aggregated rate is much smaller than the rates for individual sub-pathways. This demonstrates an important issue: we don't want to monitor (possibly) hundreds of separate performance indicators, but at the same time, aggregating sub-pathways to the pathway level loses a large amount of information; the assumed variability within individual sub-pathways is *smoothed over*. Furthermore, aggregation may result in a change in the estimated leakage rate. This is due to the combination of varying volumes passing through alternative modes of inspection; those with larger volumes will contribute more to the aggregated indicator, and consequently have higher impact on the estimated leakage rate.

A framework for monitoring performance should provide the ability to respond to both component pathways, as well as the aggregate pathway—i.e. it is a combination of the two figures. The next section proposes a framework that will permit such a general solution.

3.2. Framework for Monitoring

In this section we will suggest a framework for monitoring the performance indicators. The framework that we develop will be based on probabilities of certain events occurring, whilst accounting for various sources of uncertainty, for example variability due natural variation in the arrival rates, as well as sampling variability.

We suggest two key features in our monitoring framework:

1. Trend comparisons to judge if an indicator is increasing or decreasing over time; and

2. Benchmark comparisons to judge if an indicator is meeting minimum performance requirements.

To judge if an indicator is increasing (decreasing) over time, we use the derived calculation of the probability that the indicator increased (decreased) between two periods of interest. The choice of whether to monitor for an increase or decrease should be based on the definition of the indicator. The approaching compliance rate, for example, should be monitored for a *decrease* between periods of time, as we would like to have warning when compliance is decreasing.

The length between the two periods of interest may also be important. For example, it may be useful that the monitoring is responsive to short term shocks to the pathway, whilst also retaining the ability to detect subtle trends over the long term: if multiple short term decreases in the approaching compliance rate occur, each one by itself may not be detected with sufficient probability to produce a warning[1]. The decrease over multiple periods might, however, be significant, and this may be desirable in a monitoring framework. To simplify the monitoring framework description and the case studies, we do not include this long term monitoring further in our discussion, but we do recommend that any adoption of this framework investigate the adoption of long term monitoring.

Benchmark comparisons also feature in our monitoring framework. Appropriate target levels for indicators (set by the department and its stakeholders) can be monitored using a similar probability framework as described above. For the case of benchmarks however, we do not consider any long term effects—we focus attention on whether the probability that the indicator is higher (or lower) than the benchmark is sufficiently large.

The monitoring framework requires cutoffs to determine the state of the pathway. As we have two features to monitor (trend and benchmark of the indicator), we require a decision matrix to assign a 'health' level of the pathway. Table 3.1 shows such a decision matrix. There are two points to note about this decision matrix: (i) we require setting cutoff probabilities for the indicators themselves, and (ii) the setting of the health rating (as determined by the combination of the indicators) needs to be determined. We tentatively suggest the decision matrix in Table 3.1 as an example; further scrutiny by the department should be made to determine the actual cutoffs and assignments as appropriate. Similarly, the probability cutoffs that determine which state the pathway is in (P_1 and P_2) along with the actual decision are parameters that are required to be set by the department, in consultation with appropriate stakeholders.

[1]This may be due to low sample size in the endpoint survey.

The labels that we have attached to the health of the pathway are arbitrary, but chosen to convey meaning at a glance to the pathway manager. For a pathway where a decreasing trend and being below a benchmark is undesirable, Table 3.1 can be understood as follows:

Acceptable The pathway has a sufficiently low probability of being below the benchmark, and/or the probability of a decreasing trend is sufficiently low. No action is required at present on this pathway.

Pay Attention There is a moderate probability of being below the benchmark and a moderate to high probability that there is a decreasing trend, or there is a high probability of being below the benchmark and a low probability of a decreasing trend. Paying attention to this pathway is recommended.

Take Action There is a high probability that the indicator is less than the benchmark, and a moderate to high probability of a decreasing trend. Managerial action is recommended for this pathway.

3.2.1. Using the Decision Matrix

We now provide a brief example on how to use the decision matrix (Table 3.1) just introduced. As in the previous section, we assume that decreasing trends of an indicator are undesirable, as is the indicator being below a given benchmark. Assume that the probability cutoffs in Table 3.1 are $P_1 = 0.9$ and $P_2 = 0.6$. After analysing our pathway, we find that the probability that the indicator is decreasing is $p_i = 0.7$, and the probability that the indicator is less than the benchmark is $p_b = 0.95$. Using Table 3.1, we see that $P_2 = 0.6 < p_i = 0.7 < P_1 = 0.9$, so we are in the second row; also $p_b = 0.95 > P_1 = 0.9$, so we are in the first column. Thus in this example, our pathway would be given the **Take Action** rating.

Table 3.1.: Decision matrix to assign the health of the pathway based on monitoring the trend and level of the indicator. This decision matrix is for an indicator that should be *high*, so that being below the benchmark, or a decreasing trend is not desirable.

		Probability (p_b) that the indicator is less than the benchmark		
		$p_b > P_1$	$P_2 \leq p_b < P_1$	$p_b < P_2$
Probability (p_i) that the indicator is decreasing	$p_i > P_1$	Take Action	Pay Attention	Acceptable
	$P_2 \leq p_i < P_1$	Take Action	Pay Attention	Acceptable
	$p_i < P_2$	Pay Attention	Acceptable	Acceptable

3.3. Statistical Modelling of Leakage Rates

In order to implement the monitoring framework detailed above, we need a method to estimate the probabilities of exceeding the benchmarks and time trends of the indicators. The most appropriate way to estimate these probabilities will be via the estimation of a statistical model.

For some combinations of item class by intervention method, the rates show considerable variability in both level (how much they change over time) and precision (how well we can estimate them, i.e. the width of the probability intervals surrounding the estimates). Due to this variability, point estimates of the performance indicators will not be sufficient for a sophisticated pathway monitoring system.

As the performance indicators we use in this report are functions of known counts and estimates of the leakage rate (Section 2), we only require a statistical model for the evolution of the leakage rate. Some advantages of taking a modelling approach for the leakage rates are as follows:

- Sparsity may be *smoothed over* within each sub-pathway. In other words, if there are data gaps (for example, no endpoint surveys conducted in a sub-pathway one year for some reason), a model will allow the leakage rate to have been predicted, conditional on data from previous years and other sub-pathways;
- Sub-pathway estimates can easily be aggregated via volume counts[2]. That is, we can still calculate an overall leakage rate, adjusted for the varying sub-pathways;
- Variability due to the estimation of leakage rates is easily propagated through to the various performance indicators of interest; and
- Probability intervals can be calculated easily for any quantity of interest in the model, for example the probability that the leakage rate is increasing. This allows decisions to be made about increases/decreases in leakage (and hence performance indicators), in an unambiguous, transparently designed monitoring system.

There are some disadvantages to modelling the leakage rates as we propose in this report: the foremost being the technical expertise required to implement such models. The particular statistical models that we suggest (see Appendix A for technical details), are not basic models that could be implemented using spreadsheets, but rather require statistical modelling software—we used R Version 3.5.0 (R Core Team, 2018), and RStan Version 2.17.3 (Stan Development Team, 2018).

There are other possibilities for monitoring leakage rates and the associated indicators derived from them. Some of these methods were discussed in Fox (2007), for example Shewhart charts and exponentially weighted smoothing average charts. These methods fall under the broader term of statistical process control, and are commonly seen in industrial applications. The methods described in Fox (2007) use traditional statistical significance testing to 'signal' when the indicator deviates from a fixed (known) trend, such as that set by a benchmark. Such approaches can be semi-automated, and simple interfaces added (e.g. through RShiny).

The application under consideration is characterised by nested indicators, and (at present) a short time series of suitable data. Much of the statistical process control

[2] Also known as post-stratification.

literature has been focused on the analysis of single, long period series. A long series is required in the traditional setting due to two phases of monitoring: Phase I is used to set appropriate control parameters, and Phase II is used for monitoring. These phases mean that analysis of multidimensional series, such as those considered in this report, become difficult to manage, due to the complex structure of the data. Furthermore, there is currently no satisfactory method for dealing with high dimensional monitoring data, and particularly aggregation (see Section 3.4) of multiple series. For a review of trends and issues in statistical process control, see the review article by Woodall and Montgomery (2014).

3.4. Aggregation of Pathway Levels

As discussed in Section 1, a pathway may be further classified into *sub-pathways* according to structural components within the pathway—for example, a pathway can be broken down by location of the entry point, the item type, and the intervention method. Because of the multiple levels on which a pathway may be monitored, the framework provided in Section 3.2 can be applied at each of these levels.

We suggest that for reporting, the lowest level of the pathways be aggregated to higher levels via post-stratification. What this means is that we combine low-level performance indicators by a weighted sum of each indicator, where the weights are the total volume of the pathway within each level. Suppose that we have 4 entry points and 5 intervention methods, and we reference each combination of entry point and intervention method by the index j. Given the known volume of the pathway within each level n_j, Equation (3.1) demonstrates how to calculate the aggregated indicator value (θ_s) for a given entry point, s.

$$\theta_s = \frac{\sum_{j \in s} n_j I_j}{\sum_{j \in s} n_j} \tag{3.1}$$

where I_j is the modelled value of the indicator in the pathway level j. Items monitored by multiple methods have to be handled as a unique class.

4. Discussion

While not shown here, our case studies revealed that assessments of the health of the pathway may differ when monitoring different indicators. We found that while overall pathways might have **Acceptable** health, individual sub-pathways can have **Take Action** ratings, for example due to a greater number of non-compliant items arriving along those sub-pathways.

For our case studies we chose the probability cutoffs to determine the health ratings via the decision matrix (Section 3.2). Whilst these are sensible cutoffs, they were chosen arbitrarily, and the department may wish to choose alternative cutoffs to better align with departmental and stakeholder values.

Some benchmarks are likely to be very difficult to set. As an example, consider approaching non-compliance volume, which is dependent on the volume of items passing through each entry point. If this volume is highly variable within and between entry points, setting a benchmark for the indicator would depend on the entry point, and would be difficult to set.

Issues with different indicators notwithstanding, it is important to note that each of them does provide different information on the pathways. For example, the leakage rate and effectiveness of profiling provide information on border activities; the approaching compliance rate and approaching non-compliance volume provide information on pre-border activities and management of offshore biosecurity.

In each of the case studies that we conducted, non-compliance was determined as any failure to meet the biosecurity regulations for that pathway. As an example, a failure of compliance due to incorrect documentation would be considered as non-compliant. Consequently, the biosecurity risk of the non-compliance is not currently accounted for in the indicators we have developed. CEBRA understands that the nature of risk for each non-compliance is recorded in databases that underpin these indicators—these risk classifications were not available for analysis at this time.

It is important to reiterate that endpoint surveys are required in order to estimate approach rates for pathways that are not fully inspected, or where we cannot assume that inspection efficacy is 100% (see Section 1). In order to attempt to make an estimate of the approach compliance rate in such pathways, we need to make an assumption about the inspection efficacy. This may depend on both the location at which the inspection is taking place, and the intervention method. Nonetheless, an estimate, or more appropriately, an assumed distribution for the effectiveness will be required; such a distribution could be made via expert elicitation, however this will be very pathway dependent, and a very intensive exercise.

For fully inspected pathways (for example, the cut flowers cargo pathway), there are two options available; the first assumes that inspection is 100% effective, such that there is no leakage. The approaching compliance rate can then be estimated using all non-compliance data. The second option is to use an estimate of effectiveness (as discussed above) to inflate the observed non-compliance to account for imperfect detection. Neither of these options is entirely satisfactory, but if monitoring is required for these pathways, one of these options will need to be adopted until such time as endpoint surveys are implemented on these pathways.

Acknowledgments

This report is a product of the Centre of Excellence for Biosecurity Risk Analysis (CEBRA). In preparing this report, the authors acknowledge the financial and other support provided by the Australian Department of Agriculture, Water and the Environment, the New Zealand Ministry for Primary Industries and the University of Melbourne.

Bibliography

Fox, David R (Mar. 2007). *Statistical Methods for Biosecurity Monitoring & Surveillance*. Tech. rep. 0605. Australian Centre of Excellence for Risk Analysis.

Hoffmann, Martina, Andrew P Robinson, and Jess Holliday (June 2016). *CEBRA Project 1501F: Performance Indicators for Border Compliance*. Tech. rep. 1501F, Phase 2 Output 4. Centre of Excellence for Biosecurity Risk Analysis.

R Core Team (2018). *R: A Language and Environment for Statistical Computing*. Vienna, Austria.

Robinson, Andrew P, Rob Cannon, and Robert Mudford (Dec. 2011). *DAFF Biosecurity Quarantine Operations Risk Return Study I*. Tech. rep. 1001I, Report 1. Australian Centre of Excellence for Risk Analysis.

Robinson, Andrew P, Robert Mudford, Kathleen Quan, Paul Sorbello, and Matthew Chisholm (May 2013). *Adoption of meaningful performance indicators for quarantine inspection performance*. Tech. rep. 1101D. Australian Centre of Excellence for Risk Analysis.

Schneider, Karen and Edith Arndt (Oct. 2019). *CEBRA Project 1607B: Evaluating the Health of Australia's Biosecurity System*. Tech. rep. 1607B. Centre of Excellence for Biosecurity Risk Analysis.

Stan Development Team (2018). *RStan: the R interface to Stan*.

Woodall, William H and Douglas C Montgomery (Jan. 2014). "Some Current Directions in the Theory and Application of Statistical Process Monitoring". In: *Journal of Commodity Science, Technology and Quality* 46.1, pp. 78–94. ISSN: 0022-4065. DOI: 10.1080/00224065.2014.11917955.

A. Statistical Models

This appendix details the type of statistical model underlying the analyses performed for this report. We develop the statistical models from the most specific to the most general. This is standard in computational statistics in which we make incremental advancements.

For the purpose of the following examples, we will assume a pathway with four entry points, four item classes, and four intervention methods.

A.1. Example 1

Example 1 uses a model within a single entry point, and considers a single item class and single intervention method. We model the leakage count as arising from a Binomial distribution, with parameter n_t the number of items inspected in the endpoint survey at time t, and p_t the leakage rate to be estimated:

$$Y_t \sim \text{Binomial}(n_t, p_t) \tag{A.1}$$

We use a dynamic (random walk) model for the progression of the leakage rate over time; in particular, we assume that the leakage rate at time t depends on the leakage rate at time $t-1$:

$$\begin{aligned} p_t &= h(\theta_t) \\ \theta_t &= \theta_{t-1} + \delta_t \end{aligned} \tag{A.2}$$

where in our application we use $h(\cdot) = \text{logit}(\cdot)$.

This formulation means that we don't have to specify the functional form of the leakage rate over time: δ_t provides the trend/innovation of each time period. To complete the specification, we provide priors for the initial ($t = 1$) leakage rate and the trend term δ_t:

$$\begin{aligned} \theta_1 &\sim \text{Normal}(0, 1) \\ \delta_t &\sim \text{Normal}(0, \sigma_\delta) \text{ for } t \in 2, \ldots, T \\ \sigma_\delta &\sim \text{Cauchy}(0, 2.5) \end{aligned}$$

A.2. Example 2

Example 2 expands upon Example 1 by using data from all four entry points, but is still restricted to items of a single item class, and inspection using a single intervention method. The outcome model is again Binomial:

$$Y_{t,j} \sim \text{Binomial}(n_{t,j}, p_{t,j}) \tag{A.3}$$

for the j^{th} entrypoint at time t.

To account for entry point specific trends (with functional form unspecified), we again use a dynamic model at the (transformed) entry point level:

$$\begin{aligned} p_{t,j} &= h(\theta_{t,j}) \\ \theta_{t,j} &= \mu + \alpha_j + \delta_{t,j} \\ \delta_{t,j} &= \delta_{t-1,j} + \eta_{t,j} \end{aligned} \tag{A.4}$$

where μ is an overall intercept and α_j a entry point specific intercept. Entry point specific time dependence is modelled through the dynamic term $\delta_{t,j}$.

Again the specification is completed with priors for the initial ($t = 1$) dynamic terms $\delta_{1,j}$, and priors for the intercept terms:

$$\begin{aligned} \mu &\sim \text{Normal}(0, 1) \\ \alpha_j &\sim \text{Normal}(0, \sigma_\alpha), \forall j \\ \delta_{1,j} &\sim \text{Normal}(0, 1), \forall t \\ \eta_{t,j} &\sim \text{Normal}(0, \sigma_{\eta_j}), \forall j, t \geq 2 \\ \sigma_\alpha, \sigma_{\eta_j} &\sim \text{half-Cauchy}(0, 2.5), \forall j \end{aligned}$$

A.3. Example 3

Example 3 expands upon Example 2 by now bringing in data from all four item classes. The outcome model is again Binomial:

$$Y_{t,j} \sim \text{Binomial}(n_{t,j}, p_{t,j}) \tag{A.5}$$

for the j^{th} entry point by item class combination at time t. Here $j = 1, \ldots, 16$ as there are four entry points, and four item classes.

For ease of explanation, we will define the entry point by item class combinations as sub-pathways (e.g. Section 1). We now account for *sub-pathway* specific trends (with functional form unspecified), by again using a dynamic model at the (transformed) sub-pathway level:

$$\begin{aligned} p_{t,j} &= h(\theta_{t,j}) \\ \theta_{t,j} &= \mu + \alpha_{f[j]} + \beta_{c[j]} + \delta_{t,j} \\ \delta_{t,j} &= \delta_{t-1,j} + \eta_{t,j} \end{aligned} \tag{A.6}$$

where μ is an overall intercept, $\alpha_{f[j]}$ is the entry point specific intercept for the j^{th} sub-pathway, and $\beta_{c[j]}$ is the item class specific intercept for the j^{th} sub-pathway. Sub-pathway specific time dependence is modelled through the dynamic term $\delta_{t,j}$.

Again the specification is completed with priors for the initial ($t = 1$) dynamic terms $\delta_{1,j}$, and priors for the intercept terms:

$$\begin{aligned} \mu &\sim \text{Normal}(0, 1) \\ \alpha_f &\sim \text{Normal}(0, \sigma_\alpha), \forall f \\ \beta_c &\sim \text{Normal}(0, \sigma_\beta), \forall c \\ \delta_{1,j} &\sim \text{Normal}(0, 1), \forall t \\ \eta_{t,j} &\sim \text{Normal}(0, \sigma_{\eta_j}), \forall j, t \geq 2 \\ \sigma_\alpha, \sigma_\beta, \sigma_{\eta_j} &\sim \text{half-Cauchy}(0, 2.5), \forall j \end{aligned}$$

A.4. Example 4

Example 4 is the natural extension to Example 3, and now brings in the full cross-classified dataset including all item classes and intervention methods. The outcome model is again Binomial:

$$Y_{t,j} \sim \text{Binomial}(n_{t,j}, p_{t,j}) \tag{A.7}$$

for the j^{th} entry point by item class by intervention method combination at time t. Here $j = 1, \ldots, 64$ as there are four entry points, four item classes, and four intervention methods.

As before (Appendix A.3), we will define the entry point by item class by intervention method combinations as sub-pathways (e.g. Section 1). Our model is similar to Equation (A.6), but now includes an intercept term for the inspection method:

$$\begin{aligned} p_{t,j} &= h(\theta_{t,j}) \\ \theta_{t,j} &= \mu + \alpha_{f[j]} + \beta_{c[j]} + \gamma_{i[j]} + \delta_{t,j} \\ \delta_{t,j} &= \delta_{t-1,j} + \eta_{t,j} \end{aligned} \tag{A.8}$$

where the parameters are as in Appendix A.3, with the addition of $\gamma_{i[j]}$ as the inspection method specific intercept for the j^{th} sub-pathway.

Again the specification is completed with priors for the initial ($t = 1$) dynamic terms $\delta_{1,j}$, and priors for the intercept terms:

$$\begin{aligned} \mu &\sim \text{Normal}(0, 1) \\ \alpha_f &\sim \text{Normal}(0, \sigma_\alpha), \forall f \\ \beta_c &\sim \text{Normal}(0, \sigma_\beta), \forall c \\ \gamma_i &\sim \text{Normal}(0, \sigma_\gamma), \forall i \\ \delta_{1,j} &\sim \text{Normal}(0, 1), \forall t \\ \eta_{t,j} &\sim \text{Normal}(0, \sigma_{\eta_j}), \forall j, t \geq 2 \\ \sigma_\alpha, \sigma_\beta, \sigma_\gamma, \sigma_{\eta_j} &\sim \text{half-Cauchy}(0, 2.5), \forall j \end{aligned}$$